

THIS WEEK



EDITORIALS

WORLD VIEW The dizzying, awe-inspiring genius of space on screen **p.313**

BAD WOLF Once upon a time biologists analysed some fairy tales **p.314**

HIGHWAYS Traffic sounds from fake road drive away birds **p.315**

Climate negotiations soldier on

As the Warsaw conference on the climate wraps up this week, there is reason for hope despite several well-publicized setbacks.

One could be forgiven for feeling gloomy about the political prospects for action on global warming this week. On 15 November, as the United Nations climate-change conference rounded out its first week in Warsaw, Japan announced a drastic scaling back of its climate commitments. On Monday of this week, just as the talks kicked into high gear, Poland's environment ministry opened its doors for a conference predicated on the idea that burning coal more efficiently will reduce greenhouse-gas emissions. And on Thursday, as the negotiations head into their final hours, Australia's House of Representatives is expected to vote on a proposal to repeal the country's carbon tax.

Nobody was expecting grand things from this year's talks in Warsaw. And in truth, it requires a certain leap of faith to hope that something truly significant will come of the next big climate summit, in Paris in 2015. But one could expect that countries would not simply give up and throw in the towel.

Japan's announcement was not entirely surprising, given the shutdown of its nuclear industry following the 2011 tsunami and the resulting nuclear disaster at Fukushima. At times it has been a struggle to keep the lights on. Regardless of the course that Japan ultimately takes with regard to nuclear power, however, the country cannot simply abdicate from its climate responsibilities. Whereas Japan had previously committed to reduce emissions to 25% below 1990 levels by 2020, its new commitment would allow emissions to rise by 3.1%. An analysis by an international team of scientists that produces the Climate Action Tracker suggests that Japan could still reduce emissions by at least 17% below 1990 levels if it simply replaced all the missing nuclear power with its current blend of fossil fuels. By this measure, Fukushima is more an excuse than a justification.

In Australia, newly elected Prime Minister Tony Abbott has promised to repeal the country's carbon tax with a nebulous new 'Direct Action Plan' that will focus on incentives rather than regulations. Australians have to wait to see Abbott's alternative plan to reduce emissions, but for now the prime minister is more concerned with ridding the country of a "toxic tax".

One of Abbott's claims is that a carbon tax would put Australia's economy at a disadvantage internationally, which would be true if nobody else acts. To avoid that is, of course, the purpose of the UN conference. Collective action is needed, both to reduce global emissions and to reassure individual countries that their pain will not be in vain.

Unfortunately, the political backdrop in Warsaw this week was no more inspiring. The Polish government hosted a parallel coal conference and has put its stamp on the 'Warsaw Communiqué', which calls for the rapid advancement of "high-efficiency low-emissions coal combustion technologies" in the battle against global warming. It should go without saying that an expansion of coal-fired power, regardless of efficiency, will not protect the climate unless coupled

with — currently unavailable — technologies that enable carbon to be economically captured and buried.

Each of these cases reflects the serious challenges ahead, but there is also reason for hope. Carbon emissions fell in the United States and Europe again last year, and the rate of growth in China dropped sharply as well. Globally, carbon emissions increased by just 1.1% in 2012 compared with an average annual growth of nearly 3% over the past decade. Researchers at the Netherlands Environmental Assessment Agency in

"The goal for Warsaw this week is not an agreement but a viable roadmap to an agreement."

The Hague, who compiled the numbers, argue that this might be the first sign of a levelling off. Although the current commitments fall well short of what will probably be needed, it is also true that most countries have stepped forward with climate plans of some sort. And whereas the focus was once solely on rich countries, which are responsible for the bulk of the historical emissions but cannot halt global

warming on their own, developing nations are now putting forward mitigation plans.

In light of the dismal record over the past two decades, this represents a kind of progress. Reducing emissions will be neither easy nor free, particularly given the need to expand basic energy services in poor countries. As the climate talks in Warsaw wrap up this week, countries must seek a framework that encourages everybody to increase their ambitions and ensures that those commitments are kept. The temptation to abandon the effort and drift back into business-as-usual will always be there. The goal for Warsaw this week is not an agreement, but a viable roadmap to an agreement. Surely that much can be achieved. ■

The new zoo

Changes to the international zoological code are to be welcomed, despite continuing dissent.

The change last year that allowed zoologists to name new species in online-only academic journals was a long time coming, so it should come as no surprise that dissent continues to rumble. Publishers of journals, including this one, are keenly aware of the complexities of nomenclature, just as they are alive to the possibilities, problems and pitfalls that might have a bearing on nomenclature in a period of rapid change. The current flight from print to electronic media might (although it is too early to say) have an effect on the dissemination of information as profound as that caused by the invention of the printing press, so it is understandable that those wedded to more

traditional modes of publication might experience feelings of anxiety.

Such anxiety seems to have prompted some taxonomists to air their concerns in print. In a paper in *Zootaxa* (A. Dubois *et al.* *Zootaxa* 375, 1–94; 2013), a number of disgruntled scientists take issue with the recent change, made by the International Commission on Zoological Nomenclature (ICZN) to the International Code of Zoological Nomenclature. On the surface, their argument concerns technicalities under which certain forms of publication might render nomenclatural acts ‘unavailable’ — that is, of no taxonomic validity. If this is indeed the case, the ICZN should take these concerns seriously with a view to amending the code to ensure that its provisions are transparent and free of contradiction. A new edition of the code is scheduled for 2018, so there is ample time for consideration.

That said, there might be more than a disinterested concern for scientific integrity at work here. A typical reader of the *Zootaxa* paper (not that there are typical readers of a 94-page work on the minutiae of nomenclature protocol) might reasonably conclude that the authors have axes to grind. Exhibits A–E: the high degree of autocitation in the *Zootaxa* paper; the admission that some of the authors were against the ICZN amendments; that they clearly feel that their opinions regarding the amendments have been disregarded; the *ad hominem* attacks on ‘wealthy’ publishers as opposed to straitened natural-history societies; and the use of emotive and occasionally intemperate language that one does not associate with the usually dry and legalistic tone of debate on this subject. (The online publisher BioMed Central, based in London, gets a particular pasting, to which it has responded; see go.nature.com/vglfig.)

One of many recommendations made in the diatribe is that journals should routinely have on their review boards those expert in the business of nomenclature — in other words, a cadre of people who are, unlike ordinary mortals, qualified to interpret the mystic strictures of the code. A typical reader is again entitled to ask whom, apart from themselves, the authors think might be suitable candidates.

The naming of species is, of course, important. There was lengthy discussion of the question of permanence, and the almost-certain

enduring nature of digital publishing, before the change to the code was made. *Nature* was in favour at the time and remains so today. Simply put, the positives outweigh the negatives. As we said in an editorial when the change was announced in September 2012: “It is a sensible move, and one that most in the field should welcome ... Proper taxonomy and a robust archive are crucial to science, and the zoologists were right to consider with care the possible negative aspects of such

a change, as well as listening to the clamour to embrace the new.” (*Nature* 489, 78; 2012).

It is unfortunate that the row could overshadow more cheering news from the world of nomenclature this week. The National University of Singapore has agreed to fund the secretariat of the ICZN for the next three years. As well as administering the code, the 26 volunteer commissioners of the ICZN arbitrate on disputes between scientists over the naming of the 15,000 or so species described and named each year.

“Given the demands on their time, the ICZN members could probably do without a reprisal of the online versus print naming debate.”

Given the demands on their time, the ICZN members could probably do without a reprisal of the online versus print naming debate — a debate, remember, that saw the farcical printing to paper of hard copies of online-only papers, which were then handed to libraries to fulfil the exact wording of the code. The *Zootaxa* authors seem unwilling, or unable, to move on. They have a semantic bee in their bonnet over the code’s requirement that species descriptions must be always “available”. When the online publishers they contacted explained that, no, they did not routinely supply paper versions of the files on the journal’s websites, the authors, rather uncharitably, deemed the information unavailable to them.

This year’s must-have Christmas present in the United Kingdom is a miniature statue of a friend or relative, produced while-you-wait by a 3D printer. The technology required to make “available” a PDF file is much simpler. But then the complainants know that perfectly well already. ■

Space spectacular

Nature doesn’t usually do film reviews, but Gravity is a true great.

In his book *An Astronaut’s Guide to Life on Earth*, Chris Hadfield, former commander of the International Space Station (ISS), takes aim at the empty optimism of self-help books. Never mind thinking positive, he says — the real benefits come from preparing for the worst. This philosophy is common and necessary in space flight, and so, during ‘contingency sims’ on the ground, NASA officials would throw a series of unexpected and unfortunate events at Hadfield and his fellow astronauts, to test their responses and to work out how they could be improved. Busy dealing with an already deadly technical threat to their lives in orbit, such as a medical emergency, the trainee spacemen and women would be told: oh, sorry, but now a fire has broken out. And by the way, you’re leaking oxygen. Hadfield says he found it oddly comforting to be sitting around a table with friends and colleagues discussing, for example, how they would dispose of his corpse if he died in space.

Such a cascade of bad luck could have inspired the script of the deserved cinematic smash hit *Gravity*. (Yes, *Nature* is late to this, but the film only arrived in UK cinemas this month.) Just about everything that could go wrong for the astronauts played by George Clooney and Sandra Bullock does go wrong, so much so — and if you hate even the mildest of spoilers, then stop reading now — that when Bullock

eventually splashes back down to Earth in a remote lake, the viewer is waiting for the two-tone soundtrack and the mechanical model shark from *Jaws* to appear stage right.

As Colin Macilwain explores in a World View this week on page 313, *Gravity* is loaded with political and scientific symbolism, some subtle and some less so. The three major space-flight powers — the United States, Russia and now China — are all represented on screen, and their differing roles in the plot say much about the status of space science back on the ground in the real world.

Macilwain also celebrates the benefits the film could have for the public perception of space science, which, he summarizes, can be indistinguishable from space exploration in the public eye. Funders and scientists have quibbled for decades over the true benefit of research conducted in orbit, especially aboard the horribly expensive ISS, but there is something glorious in the fact that it is there at all.

The best stories are true, they say, and even the most spectacular film is unlikely to enthrall and enthuse a generation like the grainy pictures from the Moon landings of July 1969. *Gravity* is a work of fiction, and ardent science-fiction fans will argue for years over how good it really is. (The Oscar meanwhile, seems to be in the bag.) With tongues somewhat in cheeks, physicists have been picking holes in the depiction of Bullock’s hair in zero gravity, and complaining about how the orbits of the space hardware seem to be aligned so conveniently for the plot.

But when you watch it, none of that matters. *Gravity* is a brilliant, dizzying, awe-inspiring and downright thrilling 90 minutes. And it will both enthuse and inspire. Go and see it on the big screen while you can. And, more importantly, take an impressionable teenager with you. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Thrill of space exploration is a universal constant

In the film *Gravity*, Sandra Bullock plays Everywoman, and reminds Colin Macilwain how inspiring science and discovery still can be.

For a small boy growing up near Glasgow in 1969, the appearance on a flickering television of a man with a Scottish Borders name taking one small step for man was, to put it mildly, a seminal moment.

Yet ever since NASA's Apollo programme ended in 1972, discussion of space travel in the United States (and Europe) has been dominated by arcane arguments about whether human space flight is really cutting-edge 'science'.

For this child in 1969, space travel, discovery and science were all much the same thing. Odd that it should take a film, the glorious *Gravity*, to remind me that they still are. And that the United States and Europe have — partly at the insistence of their scientific communities — dropped ambitions for human space flight and surrendered the field to China and India. I do not lament the surrender:

I merely point out that, despite protestations to the contrary, it can lead only to the eclipsing of US leadership in global science and technology.

Gravity's plot carries a simple metaphor for the passing of the space-travel torch from US grasp. (Warning: some mild spoilers ahead.)

In a career-defining performance, Sandra Bullock plays Everyman and Everywoman, a fusion of determination and uncertainty, carrying all of our doubts and fears into orbit. Early on, the International Space Station is struck by debris and we see torn and bedraggled stars-and-stripes parachutes as the station disintegrates. Later, salvation is delivered by a Chinese re-entry pod, which returns to Earth beneath billowing parachutes adorned in a strangely neutral, red-white-and-blue livery. The studio clearly felt that a climatic scene featuring the deep-red flag of the People's Republic of China would be more than a US audience could bear. But the point is clear — wherever the movie is viewed. (Filmed and painstakingly computer-simulated over many months by director Alfonso Cuarón, *Gravity* seems almost as though it was shot on location.)

It was touching to see the old space station starring in a film: I know it so well. Back in 1984, at the commencement of the space-station project, an editorial in this journal called for its cancellation (see *Nature* 307, 1–2; 1984). Five years later, I crawled through a full-scale plywood model of it in Huntsville, Alabama. I was on Capitol Hill in Washington DC when the US House of Representatives came within a single vote of pulling the plug in 1993.

At that time, the House Committee on Science had a slogan on the wall: 'Where there is no vision, the people perish.' But the space station was not visionary enough: it was a form of retreat. The United States and its partners built a space station on their way down from space; China will do so on the way up. The

political logic is inexorable. A crewed space programme will unite and galvanize the country's people. If you have the public money — as China clearly does — there is no more obvious priority.

And the effort will bring rewards. Space flight is not like gene-sequencing or wafer fabrication, which almost anyone can do if they buy the machinery. Rocket science is, after all, rocket science — it is hard, demanding and can elude even the most technologically savvy nations, as Japan's persistent inability to master it demonstrates.

It is said that Apollo yielded only the non-stick frying pan, but that misses the point. I was in Huntsville in 1989 to visit Intergraph, a NASA spin-off computer company that developed RISC (reduced-instruction-set computing) microchips. These begat SPARC microprocessors, cheap Unix workstations and modern computer graphics. Ultimately, the

masterful, computer-generated imagery for which *Gravity* is being acclaimed was set in motion, in part, by the Apollo programme itself.

That effort was collective, yet the United States' self-narrative holds that innovation comes from individuals, including the robber barons in chinos celebrated in lame films such as 2010's *The Social Network*, which depicted the genesis of Facebook. The scientists and engineers of the Apollo programme had no public profile, earned no performance bonuses, and their crucial role in driving innovation, especially in computing and materials, has not been adequately acknowledged.

Successful as it was, NASA, in its prime, was seriously flawed. Its astronaut corps was all male and almost all white, as musician Gil Scott-Heron ruefully observed in his superb 1970 number, *Whitey on the Moon*. NASA sent up its

first African American astronaut, Guion Bluford, and its first female one, Sally Ride, only in 1983. Russia put a woman in space in 1963; China did so last year, nine years after its first man.

Many scientific missions can inspire true believers. Out once with the *Nature* staff in San Francisco, California, I met a couple of under-employed stoners, deeply in love, who earnestly informed us of the human-genome posters on the ceiling of their bedroom. The quest for the Higgs boson appeals to nerds the world over. But a scientific mission that captures the imagination of everyone is a rare and precious thing.

Just last week, I heard a three-year-old boy on an Edinburgh bus announce to his mum: "Ah want tae go tae the Moon." She lied: "You could be an astronaut!" The next person on the Moon won't have a Scottish name like Neil Armstrong's, and may not even be a man. They will, however, inspire the sort of awe whose encapsulation in *Gravity* will surely win Bullock and Cuarón their 2014 Oscars. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@googlemail.com

A SCIENTIFIC
MISSION THAT
CAPTURES THE
IMAGINATION
OF EVERYONE IS A
RARE AND
PRECIOUS
THING.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/mtcfmt

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

LAB METHODS

A strict diet for *Drosophila*

Lab fruitflies may soon all face the same limited menu.

Matthew Piper and his colleagues at University College London have developed a synthetic foodstuff for fruitflies (*Drosophila melanogaster*) that is made up of precise amounts of amino acids, vitamins and sugars that the insects need.

Feeding diverse foods to flies, as is common in labs, can drastically change their metabolism, but giving a standard food to all lab flies would ensure that it does not influence experimental results.

The researchers note that flies raised on the synthetic food grow more slowly and are less fertile than those fed a mix of sugar and yeast, suggesting that there are improvements to be made.

Nature Meth. <http://dx.doi.org/10.1038/nmeth.2731> (2013)

QUANTUM PHYSICS

A record-breaking quantum bit

Physicists have stored a quantum bit of information at room temperature for more than 39 minutes, smashing the previous record of just 2 seconds.

Mike Thewalt at Simon Fraser University in Burnaby, Canada, and his colleagues stored the bit in the nuclear spins of ionized phosphorus atoms embedded in a highly enriched silicon crystal, using optical and radio-frequency light to encode and read out the bit.

The next step is to find a reliable way to connect the nuclear spin state memory to the electronic spin states of atoms, which are more likely to be used in quantum computer



MARY EVANS PICTURE LIBRARY

CULTURAL ANTHROPOLOGY

Biology tool uncloaks folk-tale evolution

Phylogenetic analysis, a method that biologists use to infer evolutionary relationships between species, can be used to trace the ancestry of folk tales such as *Little Red Riding Hood*.

Anthropologists have struggled to find ways to group similar tales from different cultures. Jamshid Tehrani at Durham University, UK, approached the problem by creating an evolutionary 'tree' similar to those used to reveal common ancestors among biological species.

Tehrani treated each of 58 variations on *Little Red Riding Hood* as a separate species and analysed 72 varying plot elements from each tale to produce a tree displaying the tales' relationships. Notably, the analysis showed that African versions of the story are closely related to another fairy tale, *The Wolf and the Kids*, whereas East Asian versions probably evolved by combining the two with local tales.

PLoS ONE 8, e78871 (2013)

processing. Storing quantum bits at room temperature would boost efforts to create a practical quantum computer. **Science** 342, 830–833 (2013)

MARINE ECOSYSTEMS

Nutrient threat of seafood farms

Increased nutrients from aquaculture could cause harmful algal blooms in decades to come.

Lex Bouwman of the PBL Netherlands Environmental Assessment Agency in

Bilthoven and his team analysed the yearly production of farmed seafood species using data from the United Nations Food and Agriculture Organization. They estimated the amounts and types of nutrients, such as nitrogen and phosphorus, that aquaculture adds to coastal areas around the world today and predicted impacts for 2050 using scenarios from the Millennium Ecosystem Assessment.

Although most nutrient input to coastal seas comes from rivers that run through farmland, inputs from

aquaculture are growing. In some Chinese provinces, for instance, more than 20% of the dissolved nutrients in coastal waters derive from seafood farming.

Environ. Res. Lett. 8, 044026 (2013)

IMMUNOLOGY

Cells that hurt rather than heal

A type of cell that normally prevents a harmful autoimmune disease can, under certain inflammatory

conditions, cause the disease in a mouse model.

A team led by Jeffrey Bluestone of the University of California, San Francisco, studied regulatory T (T_{reg}) cells in a mouse model of an autoimmune disease in which the body attacks its own nerve tissue.

T_{reg} cells expressing the *Foxp3* gene normally act to suppress these harmful immune attacks. However, during the inflammatory response, a subset of the T_{reg} cells expressed lower levels of *Foxp3* and higher levels of proteins called cytokines.

These unstable T_{reg} cells were predominantly present in the antigen-specific T_{reg} compartment and induced anti-self immune reactions when transplanted into other mice. However, treating the T_{reg} cells with the anti-inflammatory cytokine interleukin 2 restored the cells' protective abilities.

Immunity 39, 949–962 (2013)

ARCHAEOLOGY

Teeth nibble away at invasion theory

Human remains from a fifth-century cemetery in Oxfordshire, UK, contradict the standard view of the Anglo-Saxon conquest of Britain.

Historical accounts suggest that Germanic invaders wiped out and replaced local populations at around that time. A team led by Susan Hughes of the Naval Facilities and Engineering Command Northwest in Silverdale, Washington, studied strontium and oxygen isotopes in teeth from the remains of 19 people. This can reveal whether a person ate and drank local foodstuffs.

Just one of the 19 samples contained isotopes indicating that the person came from continental Europe. The others were longtime locals, supporting the idea that Anglo-Saxons merged gradually into the region. *J. Arch. Sci.* <http://doi.org/p4j> (2013)



ANIMAL BEHAVIOUR

Phantom road frightens birds

Why did the bird not cross the road? Noise, it seems, forms at least part of the explanation.

Christopher McClure, Jesse Barber and their colleagues at Boise State University in Idaho created a 'phantom road' to test the effects of traffic noise without any actual cars or disruptions in the visual landscape.

The authors played continuous traffic sounds through speakers spaced evenly along a 0.5-kilometre ridge for four days, followed by four days of silence. They monitored multiple sites along the fake road and in a control area every morning for 7.5 weeks.

When recordings played, the number of birds along the road declined by more than one-quarter. Two species, the cedar waxwing (*Bombycilla cedrorum*; pictured) and yellow warbler (*Setophaga petechia*), avoided the noisy road almost completely.

Proc. R. Soc. B 280, 20132290 (2013)

EVOLUTION

Fish babies bigger in toxic waters

Live-bearing fish in sulphur-rich springs give birth to fewer, larger young than counterparts in non-toxic waters.

Rüdiger Riesch at the University of Sheffield, UK, and his colleagues studied nine species of fish, including the guppy (*Poecilia reticulata*), which have independently flourished in sulphur springs in the United States, the Caribbean and South America.

The researchers show that

COMMUNITY CHOICE

The most viewed papers in science

AUTOIMMUNITY

Wheat not to blame for coeliac rise

HIGHLY READ
on pubs.acs.org
14 Oct–13 Nov

The increase in cases of coeliac disease over the past 50 years or so cannot be pinned on the increasing gluten content of wheat, according to an analysis of varieties of the crop going back to the 1920s.

Some researchers have pegged the rise in the disease — an immune response to the wheat protein gluten — to modern varieties of wheat bred to contain more protein. Donald Kasarda of the Western Regional Research Center in Albany, California, compiled data on the amount of protein in US wheat crops over the past century.

Although Kasarda's analysis showed that the protein level in wheat remained largely unchanged, he did find that people now consume more wheat and foods containing gluten as an additive. This, he suggests, could account for the increase in coeliac disease since the mid-twentieth century.

J. Agric. Food Chem. 61, 1155–1159 (2013)

the toxic waters do not directly damage fish fertility. Instead, parents have fewer offspring as an inevitable trade-off of investing their energy in producing larger offspring, which can more easily detoxify hydrogen sulphide gas.

The discovery illustrates a widespread pattern of predictable evolution, they say. *Ecol. Lett.* <http://doi.org/p5h> (2013)

ZOOLOGY

Sex messes with a sea slug's head

A tiny sea slug found on Australia's Great Barrier Reef stabs its sexual partners through the head with a specialized probe, apparently to inject secretions that influence its partners' behaviour after mating.

Rolanda Lange of Monash University in Melbourne, Australia, and her colleagues observed 16 matings between 20 individuals of a newly discovered sea slug (*Siphopteron* sp.) that has a two-part penis. In all cases, seconds after the animal had inserted its penile bulb into a sexual partner to transfer

sperm, it stabbed the other part — a specialized needle-like structure (pictured) — into the head of its mate.

Related sea-slug species are known to inject prostate secretions in a similar manner, but not to the head. The authors suggest that this species is targeting the neural ganglia near the injection site, and that the secretions manipulate the behaviour of the sperm receiver.

Proc. R. Soc. B <http://doi.org/p33> (2013)



NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Fukushima fuel

Workers in Japan have taken the first steps towards fully decommissioning the stricken Fukushima Daiichi nuclear power plant. On 18 November, the Tokyo Power and Electric Corporation began the delicate task of removing fuel rods from a damaged reactor building. Although the unit was offline during the disastrous March 2011 earthquake, falling debris from the accident had made it difficult to transfer spent fuel kept in the building to permanent storage.

MAVEN launch

NASA's Mars Atmosphere and Volatile Evolution (MAVEN) spacecraft is on its way to study the upper atmosphere of the red planet. The mission launched on 18 November from Cape Canaveral, Florida, and will explore how atoms escape from the Martian atmosphere. MAVEN should reach its destination next September; once there, it will carry out a one-year nominal mission (see *Nature* **503**, 178; 2013).

POLICY

Japan emissions

Japan has scaled back its commitment to reduce greenhouse-gas emissions, according to news reports on 15 November. The country had previously pledged to lower emissions by 25% by 2020 compared with 1990 levels. But the new commitment — a 3.8% decrease over 2005 levels — would set Japan's emissions targets at 3.1% above 1990 baselines. The news comes as United Nations climate talks are under way in Warsaw. See *Nature* **503**, 174–175 (2013) and page 311 for more.



PHILIPPE MASCIET/MASTERFILMS/AIRBUS

Giant ash cloud tests sensor for aircraft

Sensors to detect volcanic ash have moved closer to widespread use on commercial airlines following flight tests involving the world's biggest artificial ash cloud (pictured). The Airborne Volcanic Object Imaging Detector (AVOID), developed by Nicarnica Aviation in Kjeller, Norway, uses infrared cameras to detect low levels of ash in the atmosphere. The test cloud was created on 30 October by spraying

particles collected from Iceland's Eyjafjallajökull volcano into the air off the west coast of France (see *Nature* **502**, 422–423; 2013). EasyJet, the UK airline carrier that helped to fund the experiment, announced on 13 November that it would mount the AVOID sensor on a number of its commercial jets by the end of next year. Volcanic ash can melt in the high temperatures of jet engines, clogging the equipment.

Brain implant

Patients with epilepsy who fail to respond to medications could benefit from a newly approved brain implant. The RNS Stimulator, made by Neuropace of Mountain View, California, received a green light from the US Food and Drug Administration on 14 November. The device is implanted in the skull and detects abnormal electrical activity in the brain, delivering corrective electrical stimulation pre-emptively to the brain areas in which epileptic seizures are thought to originate.

Malaria strategy

Researchers should aim to develop malaria vaccines by 2030 that can reduce the disease by 75%, the World

Health Organization said on 14 November in its updated Malaria Vaccine Technology Roadmap. The original 2006 roadmap had called for a malaria vaccine with an efficacy of 50% against severe disease and death — a target that seems unlikely to be met (see *Nature* **502**, 271–272; 2013). To accelerate progress, the revised plan recommends rapid assessment of new candidate vaccines using controlled studies in humans.

Heart health

Long-awaited clinical guidelines released on 12 November could change how physicians tackle cholesterol. The guidelines, issued by the American Heart Association and the American

College of Cardiology, advocate treating patients on the basis of their risk of cardiovascular disease, rather than trying to reduce 'bad' cholesterol (made up of low-density lipoprotein) to specific target levels, as had been previously recommended. See *Nature* **494**, 410–411 (2013) and go.nature.com/zxikwx for more.

Biofuel rules

The US Environmental Protection Agency proposed reducing requirements for the use of biofuels on 15 November, citing technical difficulties in meeting the current standards. The proposal would require that biofuels make up 9.2% of the US transportation fuel supply in 2014, down from 9.74%

FRANKLIN ORR in 2013. The requirement for advanced biofuels, which must reduce greenhouse-gas emissions by at least half, would drop from 1.62% to 1.33%. The rule is projected to reduce maize (corn) ethanol consumption by 3 billion litres next year, compared with 2013.

PEOPLE

Science educator

Microbiologist Ann Reid will be the new head of the US National Center for Science Education (NCSE) in Oakland, California. The non-profit organization campaigns against the teaching of creationism and climate-change denial in schools. Reid, formerly director of the American Academy of Microbiology in Washington DC, will replace retiring NCSE director Eugenie Scott, who has led the organization for 27 years (see *Nature* **497**, 287–288; 2013).

US energy nominees

Chemical engineer Franklin Orr (pictured) has been tapped by US President Barack Obama to serve as the undersecretary for science at the Department of Energy. Orr, a researcher at Stanford University in California, would replace Steven Koonin as chief scientific adviser to US energy secretary Ernest Moniz, and would oversee



the department's research programmes. Meanwhile, Marc Kastner, a physicist at the Massachusetts Institute of Technology in Cambridge, was nominated to head the department's Office of Science. Both picks must be confirmed by the Senate. See go.nature.com/geeup9 for more.

RESEARCH

Acidic waters

Oceans are acidifying at an “unprecedented rate”, with potentially dire consequences for humans, according to a report released on 14 November from the Third Symposium on the Ocean in a High-CO₂ World. The assessment reviews current science on ocean acidification and warns that many species will fare worse in the future. Oceans will be less able to take up atmospheric CO₂, decreasing their capacity to

moderate climate change, the report says. Shellfish harvests will probably decline and coral reefs will be lost, it adds. See go.nature.com/cjeog for more.

Heat tracking

This year is on track to become the seventh-warmest since global climate records began in 1850, according to a preliminary assessment released on 13 November by the World Meteorological Organization. See go.nature.com/apxln for more.

Asian unicorn

A rare antelope-like animal called the saola has been caught on film for the first time in 15 years. The conservation group WWF snapped the photos using a camera trap in Vietnam, and announced the finding on 12 November. The saola (*Pseudoryx nghetinhensis*), sometimes nicknamed the Asian unicorn, is critically endangered, and probably only a few hundred remain. See *Nature* **484**, 424–425 (2012) and go.nature.com/yerkoh for more.

FUNDING

Broad investment

American philanthropists Eli and Edythe Broad announced on 14 November a US\$100-million investment to continue funding biomedical research

COMING UP

22 NOVEMBER

The European Space Agency is scheduled to launch Swarm, a constellation of satellites that will study Earth's magnetic field for four years.

go.nature.com/rxaaur

24–27 NOVEMBER

Science for global sustainable development is the theme of the sixth World Science Forum, to be held in Rio de Janeiro, Brazil. Highlights include biodiversity, water security and bioenergy.

go.nature.com/cxmbqf

at the Broad Institute in Cambridge, Massachusetts. The centre, which was founded in 2004, supports collaborations between scientists at the Massachusetts Institute of Technology and Harvard University. The Broad launched the institute with an initial \$100-million gift, and have already contributed a further \$500 million (see *Nature* **455**, 149; 2008).

BUSINESS

Breakthrough drug

The US Food and Drug Administration (FDA) approved on 13 November a ‘breakthrough therapy’ to treat a rare blood cancer called mantle-cell lymphoma. Ibrutinib, developed by Pharmacyclics of Sunnyvale, California, is only the second drug to be approved under the FDA's Breakthrough Therapy Designation programme — a pipeline launched last year to fast-track regulatory approval of particularly promising treatments. See go.nature.com/w5xfjo for more.

► NATURE.COM

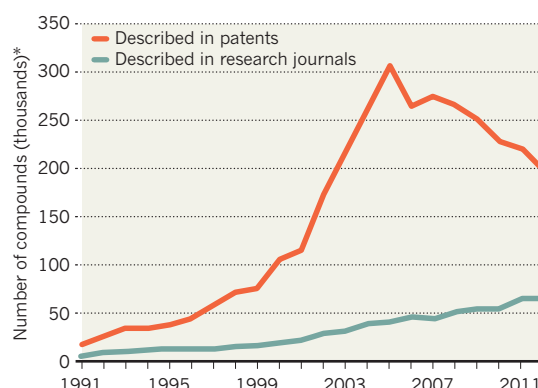
For daily news updates see:
www.nature.com/news

TREND WATCH

The number of potential drug leads disclosed in patents each year has plummeted over the past seven years. But potentially bioactive molecules described in research journals are still rising, according to a data-mining study of molecular structures in more than 140,000 journal articles and patents (C. Southan *et al.* *PLoS ONE* **8**, e77142; 2013). The researchers suggest that job cuts and mergers among pharmaceutical companies may be behind the fall in global output.

PATENT CHEMISTRY ON THE SLIDE

The number of chemical compounds linked to protein targets being disclosed in patents is on the decline.



*Linked to human protein targets (pattern similar for all species).

NEWS IN FOCUS

CONSERVATION Are fences good or bad for lions? The debate roars **p.322**

DISASTERS Scientists set to glean lessons from typhoon Haiyan **p.324**

SPACE Hopes high for mission to finally tackle 1970s X-ray enigma **p.325**

GRAPHENE Europe's quest to harness a supermaterial **p.327**



XINHUA/PHOTO SHOT



Mexican President Enrique Peña Nieto has resolved to improve the country's standing in science.

POLICY

Mexico bolsters science funding

President aims to boost spending and reform research laws.

BY LAURA VARGAS-PARADA AND ERIK VANCE

It has the world's 11th-biggest economy and is home to the largest university in the Western Hemisphere. But for all that, Mexico has had surprisingly little influence on global science output and innovation. Its annual rates of patents and spending on science lie below those of Brazil, its chief Latin American competitor.

But when Enrique Peña Nieto was sworn in as president last December, he promised to grease the rusty wheels of Mexico's science and technology infrastructure. And one year in, he has started to deliver.

On 13 November, the Mexican Congress approved a 20% rise in the 2014 budget of the National Council for Science and Technology (CONACYT) in Mexico City, the country's

main research funding agency. Congress increased the country's overall science budget by 12%, to 82 billion pesos (US\$6.3 billion). Peña Nieto is also pushing several other pieces of legislation through the pipeline: an intellectual-property bill that would allow researchers and universities to commercialize their publicly funded work; a bill that would reform the academic retirement system and encourage talented young researchers to stay in Mexico; and tax breaks that could incentivize private investment in research and development.

As head of the party that dominates both houses of Congress, Peña Nieto is in a strong position. By the end of his six-year term, he wants Mexico's combined public and private spending on science and technology to rise to at least 1% of gross domestic product (GDP). For years, the country's spending has

languished at a level of about 0.4%. By comparison, Brazil spends more than 1% of its GDP on science and technology and the United States almost 3% (see 'Peso power').

"Since the campaign, as a president-elect, and finally when he took office, President Peña Nieto has made clear that science, technology and innovation would be central for economic development and social well-being," says Gabriela Dutrénit, head of the Scientific and Technological Advisory Forum, a prominent independent science think tank in Mexico City. She says that last week's budget would put the nation on track to reach spending of almost 0.55% of GDP in 2014 — a pace not quite fast enough to reach 1% in 2018, but still an important first step.

One of the most important signs of change might not be a policy, but the creation of a scientific institution. Within days of taking office, Peña Nieto tweeted that he would form an executive-branch office modelled on the US Office of Science and Technology Policy, to advise the president on scientific matters, coordinate policies between science ministries and propose legal reforms. In April, Peña Nieto chose as its head Francisco Bolívar Zapata, a former president of the Mexican Academy of Sciences and a biotechnologist who helped to start the company Genentech in San Francisco, California, and has worked on engineering bacteria to produce human insulin. In an interview with *Nature*, Bolívar says that he pushed for science to be included in Mexico's 2013–18 development plan. About 30% of that plan's 800-plus lines of action are related directly or indirectly to science, he says.

Congress is now turning to an intellectual-property bill. Patents in Mexico are complicated and expensive, and scientists working in public research centres cannot make money on them. But a proposed law modelled on the 1980 US Bayh–Dole Act would give researchers and universities the right to develop commercial products based on publicly funded research. Rubén Félix Hays, a member of Congress who presides over the science and technology committee, says that a group of legislators is actively working on the bill.

Tony Payan, who specializes in Mexican studies at Rice University in Houston, Texas, says that the patent reform would be a good first step. But he adds that this needs to be followed by changes in the judicial system, which rarely enforces laws on intellectual-property ►

► rights. “If somebody violated your patent and you found out that they are marketing a product that is very similar to the one that you hold a patent to, what court would you go to? Where would you sue?” asks Payan.

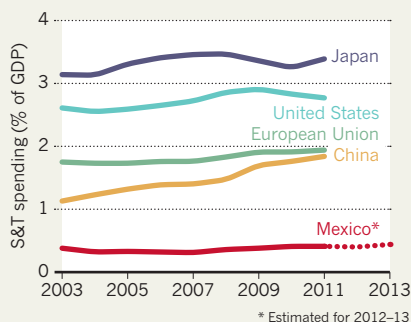
Another problem is Mexico’s massive brain drain. The reason why many scientists leave the country is clear enough: jobs are hard to come by. Scientists tend to stay in their jobs for as long as possible, because leaving means giving up most of their salaries. “We don’t retire,” says Bolívar. “We don’t free some positions for young scientists.”

To that end, the Congress is working on a bill that would boost pensions for retiring researchers. Félix says that the process might take some time. “We don’t want to harm the rights of the professors, but we need a reform so the new generations can have access and refresh the system,” he says. CONACYT also plans to create 500 new science jobs for young researchers in 2014. The jobs will be in fields such as climate research, disaster mitigation, diabetes and plant genetics. Bolívar says that the government plans to follow the first batch with 500 more each year.

However, Payan says that it will take a long time to achieve a culture of innovation, and that merely replacing the old with the young will not suffice. “You can retire a bunch of guys and you can put in the new people to work,” he says.

PESO POWER

Mexico is trying to reach a goal of spending 1% of gross domestic product (GDP) on science and technology (S&T).



“That doesn’t mean they’re going to innovate.”

By itself, a boost in public spending on science will not be enough for Mexico to achieve its goal of 1% of GDP; private investment is also needed. To encourage this, Bolívar has enlisted the help of Enrique Cabrero Mendoza, who was appointed head of CONACYT in January. A competitiveness expert at the Center for Research and Teaching in Economics in Mexico City, Cabrero has identified cities throughout the country that are ripe for investment as technology hubs. The government wants to offer corporate tax breaks

to encourage investment in these hubs — although tax breaks have been controversial in the past because they have been abused.

Peña Nieto has started to run into opposition to this and other parts of his agenda. Major reforms in education and energy policy — such as compulsory teacher evaluations and opening up the state-owned oil company to private investment — have sparked large protests in the streets, supported by powerful unions.

Even if Peña Nieto has trouble enacting all of his research agenda, his symbolic actions have already impressed Dutrénit. She points out, for instance, that in September the president reconvened a high-level scientific advisory body — the General Council for Scientific Research, Technological Development and Innovation — headed by himself and nine ministers, as well as officials from CONACYT, universities and businesses. The council was created in 2002 to help set national science and innovation policy. It is supposed to meet twice yearly, but had met only three times in the past ten years.

The government’s renewed focus on science is spurring a sense of responsibility among Mexico’s scientific elite, Dutrénit adds. “We are not only asking for increases in public and private investment,” she says. “We also have to answer for those investments.” ■

SOURCE: OECD/CONACYT

JOURNALS

PLOS profits prompt revamp

Incoming boss plans peer-review shake-up at Public Library of Science.

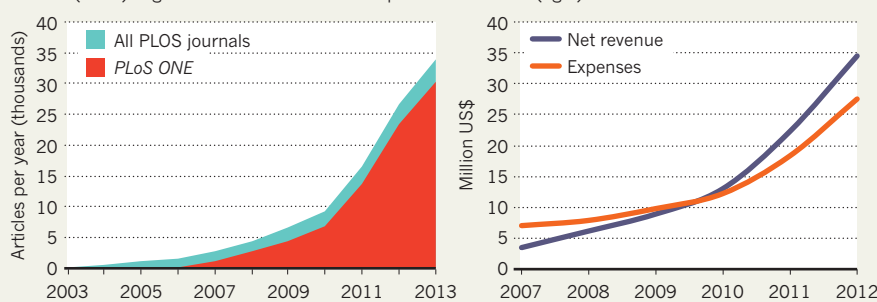
BY RICHARD VAN NOORDEN

The Public Library of Science (PLOS) is not accustomed to having spare cash. Founded by scientists in 2000 as a grassroots organization advocating open scholarly communication, PLOS reinvented itself as an open-access journal publisher in 2003 with the help of philanthropic grants. It has spent much of the decade since then “skating on thin financial ice”, in the words of co-founder and board member Michael Eisen, a geneticist at the University of California, Berkeley.

Now PLOS is part of the establishment: open-access publishing has entered the mainstream. The non-profit operation, based in San Francisco, California, broke even for the first time in 2010; in 2012, it reported a surplus of US\$7 million on net revenues of \$34.5 million. Its cash-generating engine is the world’s largest journal, *PLoS ONE*, which is on course to publish more than 30,000 articles this year (see ‘World’s largest journal’), although its growth rate shows

WORLD’S LARGEST JOURNAL

By quickly expanding the size of its megajournal *PLoS ONE* (left), the Public Library of Science (PLOS) began to see revenues exceed expenses from 2010 (right).



signs of slowing. The ‘megajournal’ business model has been mimicked by many others.

PLOS is now seeking a new vision to match its new profitability. In May, it announced the departure of chief executive Peter Jerram and the recruitment of his replacement, Elizabeth Marincola. She says that the future of science

publishing is not in branded, highly selective titles. Instead, she sees a world in which article metrics and community judgements help the cream of research to rise to the top. “The packaging of a journal will become less and less important,” she says.

That idea is the opposite of an open-access

SOURCE: PLOS

competitor of which Marincola was previously chair: *eLife*, an elite journal funded with more than £15 million (US\$24 million) from the Wellcome Trust in London, the Max Planck Society in Munich, Germany, and the Howard Hughes Medical Institute in Chevy Chase, Maryland. “Their appeal is that there is quality inferred from the brand,” notes Marincola.

“We are working to evolve all of PLOS towards a world where papers are only rejected when they are scientifically invalid,” says Eisen. *PLoS ONE* already adopts that approach, but the publisher has six more-selective journals, including *PLoS Medicine* and *PLoS Biology*. Marincola will not be drawn on whether these might become less selective, although she says that in the longer term, “we would like very much to be able to move away from our current system of peer review altogether”. The organization’s research arm, PLOS Labs, founded this year, aims to develop and test concepts for peer review after papers have published.

Others have different priorities. “One of the areas I would love to see PLOS push is doing open science cheaper,” says Jonathan Eisen, Michael’s brother and an evolutionary biologist who is on the editorial board of *PLoS Computational Biology*. Reducing the \$1,350 author fee for its lowest-cost journal, *PLoS ONE*, also makes sense tactically, says Joseph Esposito, a publishing consultant based in New York City, because it will make it harder for new entrants to break into the megajournal market. “Right now, PLOS is by far the scale leader. They should play that card now and play it aggressively,” he says. But Marincola says that PLOS has not raised its prices in four years, and waived about \$4.3 million in publishing fees last year.

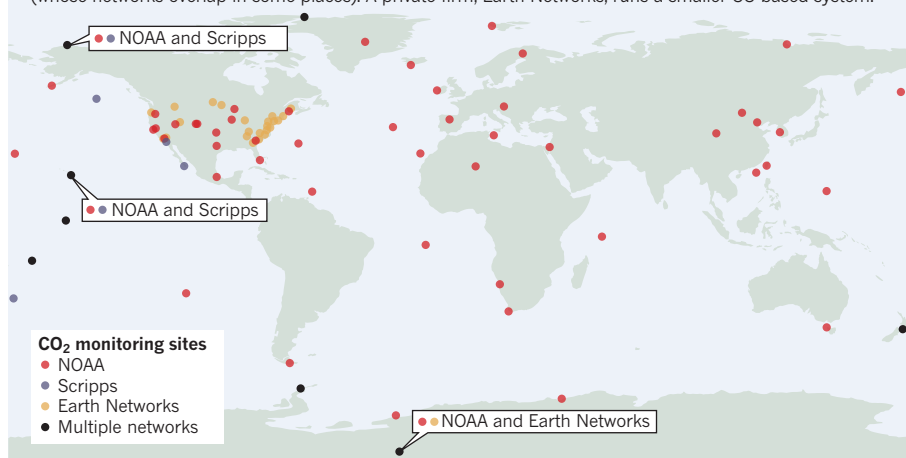
Making everything as cheap as possible is not a pressing priority, agrees Damian Pattinson, editorial director of *PLoS ONE*. Like Marincola, he thinks the immediate focus will be on iterative improvements to the publishing process. “For years, journals have got away with treating authors like scum,” he says. Open access focuses publishers’ minds on giving authors services they value, such as faster turnaround, better websites and metrics on who is viewing articles, he adds.

To Michael Eisen, some of the most visible manifestations of innovation are with other publishers — such as F1000 Research in London, which already uses open peer review after papers are published. “They are doing lots of things that PLOS should have done five years ago,” he says. “PLOS has created the landscape that has enabled others to flourish, which is great. The question is, how can it continue to be innovative?” ■

“The packaging of a journal will become less and less important.”

GREENHOUSE GRID

Scientists who monitor the build-up of carbon dioxide in the atmosphere depend on data collected by the Scripps Institution of Oceanography and the US National Oceanic Atmospheric Administration (whose networks overlap in some places). A private firm, Earth Networks, runs a smaller US-based system.



SOURCE: SCRIPPS INST./NOAA/EARTH NETWORKS

CLIMATE CHANGE

Budget crunch hits Keeling’s curves

Scientist struggles to maintain long-standing carbon dioxide record and more recent atmospheric-oxygen monitor.

BY JEFF TOLLEFSON

Late last month, officials at California’s Scripps Institution of Oceanography turned to Twitter seeking donations to maintain the iconic ‘Keeling curve’, a 55-year record of rising atmospheric carbon dioxide levels. An appeal for funds launched in July had attracted only a few small contributions, not nearly enough to keep the programme going.

Scripps geochemist Ralph Keeling, who took over the CO₂ measurements started by his father Charles, is neither surprised nor disappointed. “That’s more a fishing expedition than anything,” he says of the nascent crowdsourcing at Scripps in La Jolla. But he is worried.

For years, he has struggled to cobble together enough cash to support the CO₂ programme and an atmospheric-oxygen record that he pioneered in 1989. Bouncing between grant programmes designed to fund short-term projects, not long-term monitoring, he has cut staff and streamlined operations to keep the records going.

But now, with his funds running dry, he wonders about the future. “Things have never been this dire before,” he says.

Much has changed since 1958, when Charles Keeling took his first CO₂ measurements atop

Mauna Loa in Hawaii. The programme he started now monitors CO₂ at 13 sites, from the South Pole to Alaska (see ‘Greenhouse grid’). The National Oceanic and Atmospheric Administration (NOAA) runs a larger network that overlaps with the Scripps system, helping both teams to ensure that their measurements are correct. These data, along with other measurements from researchers around the world, flow into models designed to study how carbon moves through the environment.

The complement to the Keeling curve is Ralph Keeling’s atmospheric-oxygen record, which NOAA does not replicate. Keeling has documented a decrease in oxygen levels that is due to fossil-fuel combustion, which uses up oxygen and releases CO₂. By accounting for both CO₂ and oxygen levels in the atmosphere, scientists have calculated that oceans and plants each absorb roughly one-quarter of humanity’s CO₂ emissions, leaving half to build up in the atmosphere.

“We expected an answer close to that, more or less, but Ralph Keeling was the first to provide the measurements,” says Pieter Tans, who heads NOAA’s carbon-cycle and greenhouse-gas group in Boulder, Colorado.

Keeling says that he received around US\$700,000 annually for the CO₂ programme through paired support from the National

► Science Foundation (NSF) and the Department of Energy (DOE) until three years ago, when the NSF halted funding. With staff cuts, he has been able to maintain operations with a budget of around \$350,000. He has also partnered with Earth Networks, an atmospheric-monitoring company based in Germantown, Maryland, which has deployed a sensor for him at Mauna Loa to reduce the costs for Scripps. His latest grant application to the DOE is pending, but in the meantime he is operating on spare funds.

The situation is murkier for the oxygen measurements, which the NSF and NOAA supported for more than two decades. The NSF pulled the plug in 2009, and Keeling's NOAA grant could run out in early January. In an effort to keep things going, Keeling says that he went back to the NSF and was assured that he would get about \$350,000 from the Division of Polar Programs this autumn. (NSF officials say that they cannot comment on pending grants.) That money would carry him into next year, but it remains unclear what will happen after that. Jim Butler, director of NOAA's Global Monitoring Division in Boulder, says that NOAA cannot simply fold Keeling's CO₂ stations into its own observations budget, given that the value of having two CO₂ networks is scientific independence. The oxygen measurements, by contrast, would fit nicely into NOAA's portfolio, Butler adds, but his division's budget has shrunk by 12% since 2011, with further cuts expected this fiscal year. Budget constraints have already forced the agency to reduce staff and shut down monitoring at ten sites.

"NOAA's budget is getting hammered, and it's increasingly difficult to fund things like Ralph's programme," Butler says. "All I can do right now is provide moral support to keep it going year by year until we come up with a plan."

For a while, it seemed that commercial interests might pick up some of the slack. Working with Scripps, Earth Networks announced plans in 2011 to deploy a global network of 100 greenhouse-gas monitoring stations. But two years later, with climate regulations on the back burner in Washington DC, the company is operating just 25 stations. "We really don't have any material customers at this point," says Earth Networks' president Bob Marshall.

Keeling has considered approaching private foundations for help, but acknowledges that atmospheric monitoring is an unusual target for philanthropy. Moreover, he says, a private donor would probably want to see evidence of stable government support before committing. "The difficulty of keeping these things going long term, even within the government, needs to be recognized," he says. ■



Lion numbers have fallen sharply in recent decades, in large part because of killing by humans.

ECOLOGY

Fences divide lion conservationists

Some say enclosures offer protection, others maintain they are a menace.

BY TRACI WATSON

Times are grim for the king of the beasts. Roughly 35,000 African lions roam the savannahs¹, down from more than 100,000 half a century ago, thanks to habitat loss, declining numbers of prey animals and killing by humans. One study estimated that fewer than 50 lions (*Panthera leo*) live in Nigeria and reported no sign of the animal in the Republic of the Congo, Ghana or Côte d'Ivoire².

Now a king-sized controversy is brewing over a proposal to shore up lion populations before it is too late. A prominent lion researcher has called for limiting conflict between humans and lions by erecting fences around reserves containing wild lions. The idea has split scientists, with those opposed to the idea arguing that fences could do more harm than good. The ensuing debate has also laid bare fundamental differences of opinion about how to preserve lions and other species,

and raised concerns that a key challenge to lion conservation — lack of funds — is being ignored while scientists trade jabs about fences.

When he began the research that kicked off the furore, Craig Packer of the University of Minnesota in St Paul, who studies lions at Tanzania's Serengeti National Park, intended to determine only the cost of conserving lions. But something more provocative emerged from his data. In work reported earlier this year in *Ecology Letters*³, he and 57 co-authors calculated lion densities at 42 African reserves and found, Packer says, that the only variables that matter for density are "dollars and fence — nothing else". He adds that "the fence has a very profound, powerful effect", because it prevents lions from preying on livestock and people, meaning fewer lions are killed in retaliation. Packer would like to see fences around even some of the largest protected areas such as Tanzania's 47,000-square-kilometre Selous Game Reserve.

But the paper triggered heated discussion,

ROY TOFT/NATIONAL GEOGRAPHIC/GETTY

both online and at meetings, leading four months later to the publication of a response signed by 55 researchers⁴. They argue that Packer's analysis is wrong to use lion population density as its sole yardstick. By that measurement, they say, a dense population of several dozen lions in a small reserve is a success, whereas a large reserve containing 600 lions is a failure. When the authors⁴ restricted their study to lion populations whose density did not exceed the land's capacity to support them and controlled for a reserve's management budget, they found no relationship between fencing and density.

That study's first author, Scott Creel of Montana State University in Bozeman, says that although fencing is beneficial at small, well-funded reserves, most of Africa's wild lions live in large reserves with modest funding. "If you build a fence and spend a lot of money, you can maintain a lot of lions within it," Creel says. "The problem is, we don't know very much about how fencing works in enormous ecosystems that have smaller budgets."

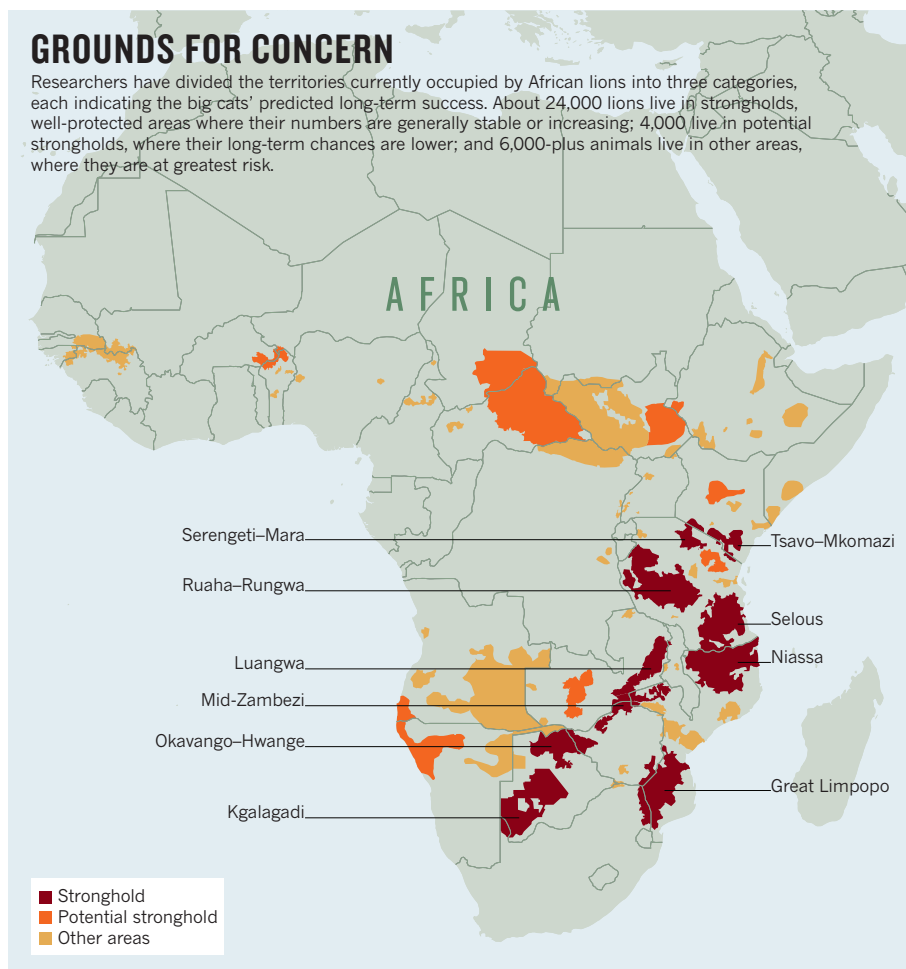
Packer's side responded with its own reanalysis⁵. Rather than eliminating the super-saturated lion populations from the equation, the researchers assigned them a density of 100%. Once again, they found the presence of fencing to be the strongest predictor of lion density, Packer says. Creel counters that the reanalysis shows no impact of fencing on population size, so it is still unclear whether fences would have a protective effect for large, natural ecosystems.

Other researchers are split over which argument is more convincing. Matt Hayward of Bangor University, UK, who co-authored a book about conservation fencing, says that both sides score points, and, in any case, the disagreement goes beyond statistics to "a very passionate philosophical debate". He adds that "some people are saying, 'Look, we don't want any fences in the landscape. We want to keep wildlife moving wherever it wants to.'"

And rightfully so, many say: ill-conceived fences could hinder animals' search for food during tough times, as well as leading to losses of wide-ranging species, such as cheetahs and wild dogs, that need big expanses of land. "Is saving lions above everything else?" asks Creel's co-author Nathalie Pettorelli of London's Institute of Zoology. "You cannot manage a landscape by looking at just one species."

GROUNDS FOR CONCERN

Researchers have divided the territories currently occupied by African lions into three categories, each indicating the big cats' predicted long-term success. About 24,000 lions live in strongholds, well-protected areas where their numbers are generally stable or increasing; 4,000 live in potential strongholds, where their long-term chances are lower; and 6,000-plus animals live in other areas, where they are at greatest risk.



Bush-meat snares made with wire stripped from fences pose another risk. These often catch and may kill lions, elephants and other species, and have been a serious problem in places such as Zambia. But Packer says that a properly built fence, although costly, would not support snaring. And he argues that the opposition's goal of maintaining open landscapes is deeply impractical in the face of Africa's burgeoning human population.

He has already tried to drum up support for fencing with African officials, and he also hopes donors with an interest in conservation projects, such as the World Bank, might fund a fence around a large reserve. Meanwhile, many of those who oppose the idea would rather see money poured into proven approaches such as law enforcement.

But the two camps also share plenty of common ground. Creel's co-authors say that fences can be effective, and Packer's allies agree fences are inappropriate for many areas.

While scientists wrangle over the issue, "lions are disappearing faster than ever," says Philipp Henschel, a lion specialist for the conservation group Panthera, who signed Creel's paper (see 'Grounds for concern'). The community should "concentrate on the one thing that both sides agree on: that effective lion conservation will require substantially more funding than is currently made available". ■

1. Riggio, J. *et al. Biodivers. Conserv.* **22**, 17–35 (2013).
2. Henschel, P. *et al. CATnews* **52**, 34–39 (2010).
3. Packer, C. *et al. Ecol. Lett.* **16**, 635–641 (2013).
4. Creel, S. *et al. Ecol. Lett.* **16**, 1413–e3 (2013).
5. Packer, C. *et al. Ecol. Lett.* **16**, 1414–e4 (2013).



**MORE
ONLINE**

TOP STORY



Octopus tentacles have 'a mind' of their own bit.ly/17wsx1z

MORE NEWS

- Cool laboratories can skew results of cancer studies in mice bit.ly/17GzfHO
- Bailout for body that oversees animal nomenclature bit.ly/17EFjeC
- Lyme pathogens can evolve to evolve bit.ly/1a9JHpS

SLIDESHOW OF THE WEEK



A behind-the-scenes look at goods from illegal trade bit.ly/1d2IUfj



A storm surge as high as 6 metres devastated hundreds of kilometres of the Philippine coast.

NATURAL DISASTERS

Haiyan prompts risk research

Geologists, engineers and social scientists are poised to swoop in before reconstruction gets under way.

BY SARAH ZHANG

When typhoon Haiyan pummelled the Philippines earlier this month with winds of more than 300 kilometres per hour, it was the most intense storm to hit land in modern history. But to truly understand how unusual a storm such as Haiyan is, scientists have to turn to the geological record. That is why Davin Wallace, who studies the traces of ancient storms at the University of Southern Mississippi in Hattiesburg, is angling to go to the Philippines in the next few weeks. He hopes to calculate how often large storms strike the Philippines by comparing coarse-grained sand deposited by Haiyan with similar layers found in metres-deep sediment cores that chart thousands of years of history.

Right now, food, shelter and sanitation are the top priorities in the Philippines, where nearly 5,000 people have died and more than 4 million have been displaced as a result of Haiyan. But in a brief window of time — after the immediate

humanitarian relief effort but before long-term rebuilding — scientists have a unique laboratory in which to gather data in fields as diverse as climatology, civil engineering and social science. Researchers who study natural disasters know that their work hinges on the misfortune of others, but they hope that the research can make future catastrophes less deadly.

For such time-sensitive research, being nimble is key. Timing is unpredictable, and everything happens fast, with just a few weeks from drafting a grant proposal to stepping off a plane into the disaster zone. “The logistics of just making everything work, that’s 80% of your time,” says Andrew Kennedy, a civil engineer at the University of Notre Dame in Indiana, who studies the threat of storm surges. “The 20% — planning for the scientific stuff — is easy by comparison.”

After Hurricane Sandy battered the coast of New Jersey in 2012, Kennedy’s team went door-to-door in one coastal town to detail the damage to more than 600 houses. He chose the

area on the basis of post-storm satellite images that showed a wide range of damage, so that he could learn why some houses were “knocked to matchsticks” whereas others were just missing a few roof shingles. Many houses had weak connections to their foundations and could not withstand the horizontal force of the storm surge. Areas not protected by high sand dunes also fared poorly. To try to understand the effect of storm surges and wave dynamics, Kennedy drops gauges onto the sea bed. Now he wants to extend his damage-prediction model, which combines ocean and building data, to the Philippines.

But with few predictable sources of grants, Kennedy knows he will have to hustle. “If it looks like something is going to hit, I call up everyone I know and ask, ‘Can you give me a little bit of money?’” he says. He has won small grants from the US Army Corps of Engineers and the US Geological Survey. The National Science Foundation has a dedicated programme, called Rapid Response Research (RAPID), which fast-tracks proposals requiring time-sensitive data collection. Although the review cycle is compressed from months to weeks, it still takes time for the cheque to arrive. After being told that RAPID would fund his Sandy research, Kennedy was on the ground within a month. But he only got the money more than a month after he returned.

Timing a research visit can be tricky in the post-disaster chaos, says Louise Comfort, a political scientist at the University of Pittsburgh in Pennsylvania, who wants to study the responses of Filipinos and relief organizations. After the earthquake in Haiti in 2010, she spent several days interviewing officials from governmental and non-governmental organizations. She found considerable distrust between Haitians and international aid organizations, which operated in English and often neglected to build local partnerships. Comfort says that arriving five weeks after a disaster is

“You as a scientist don’t want to interfere with the bigger picture going on.”

about right for a balance between letting relief workers do their jobs, and interviewing while experiences are still fresh in people’s minds. “I have gone immediately after the disaster,” she says. “When that’s the case, you get to see things as they’re happening but it’s very difficult to interrupt.”

For Wallace, time is also of the essence. He has to get his sampling done before reconstruction alters the sediment deposits that have washed up on shore. “Obviously there’s a fine balance and I for one know this balance,” he says, recalling his experience as an undergraduate in New Orleans when Hurricane Katrina devastated the city. “You as a scientist don’t want to interfere with the bigger picture going on.” ■

SPACE

Astronomers call for X-ray polarimeter

NASA explorer programme raises hopes of mapping directional light from pulsars and black holes.

BY EUGENIE SAMUEL REICH

When astronomers in the 1970s first observed polarized X-rays streaming from the Crab Nebula, they opened a new window on the Universe. Since then, scientists have proposed multiple space missions to explore other sources of these rays, such as pulsars, black holes and supernova remnants. Three spacecraft nearly flew — but time and again, space agencies cancelled or passed over these polarimetry missions. So far, the Crab Nebula, a powerful pulsar shrouded by a supernova remnant, is the only source of polarized X-rays that has been mapped.

Now, a competition for a small, US\$125-million space mission, announced by NASA on 12 November, has X-ray astronomers eagerly bidding for the chance to launch a dedicated polarimeter that would map hundreds or thousands of sources.

“The situation for polarimetry is so bad, even a small mission would be a breakthrough,” says Enrico Costa, an astrophysicist at the Italian National Institute for Astrophysics in Rome.

Polarization is the oscillation of electromagnetic waves in a particular orientation. It typically encodes geometric information about the direction in which photons are generated. It might be used, for example, to determine the axis of a spinning pulsar.

But these are not easy measurements to make. X-ray astronomy is already a challenge — Earth’s atmosphere absorbs the rays, and so observatories must orbit in space. Capturing X-ray polarization is harder still. The 1970s detectors relied on antennas made of graphite crystals, with atoms separated at the nanometre scale of X-ray wavelengths. Modern detectors use a container of gas: incoming X-rays ionize the gas atoms, which kick off electrons in the direction of the polarization. These electrons are then tracked.

But efficiency is still a problem. It takes about 100 times longer to measure the polarization of an X-ray source than it does to measure its energy and brightness. When projects have been scaled down to save money, polarimeters tend to be the first to go, says Martin Weisskopf, an astrophysicist at NASA’s Marshall Space Flight Center in Huntsville,

Alabama, who made the measurements on the Crab Nebula (M. C. Weisskopf *et al. Astrophys. J.* **208**, L125–L128; 1976).

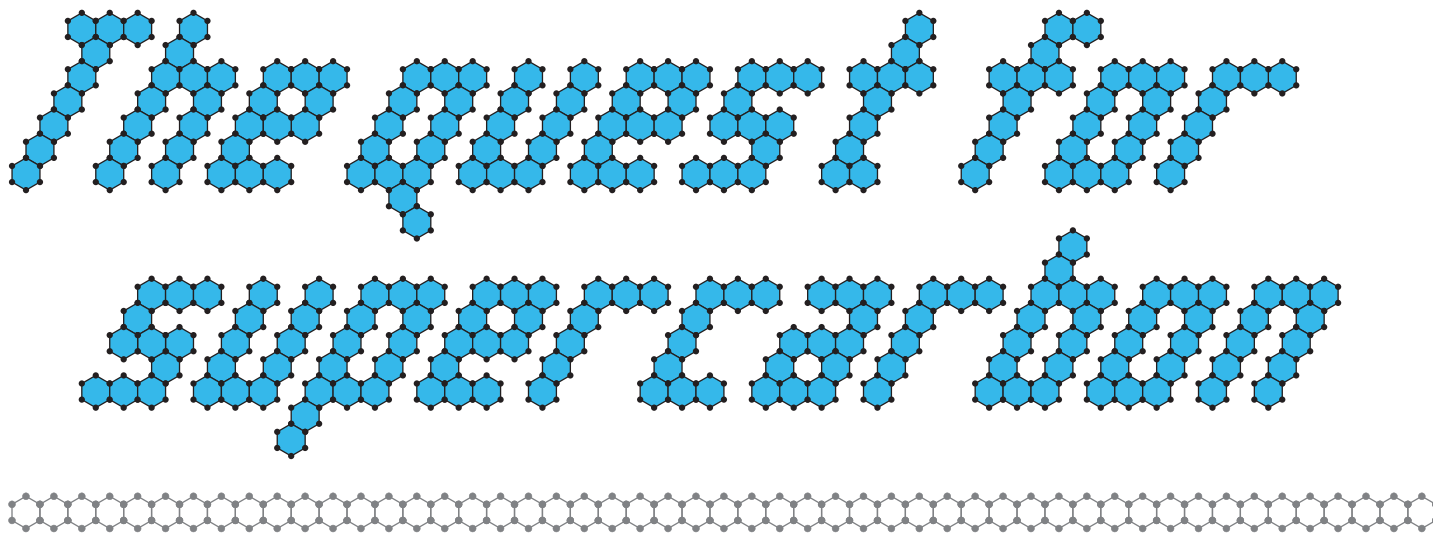
The NASA announcement might change that. Polarimetry enthusiasts are hopeful because in 2009, NASA selected a polarimeter mission called GEMS (Gravity and Extreme Magnetism) for development. It was cancelled in 2012 after managers judged it was likely to run over its \$105-million budget. But the GEMS team has since tested detectors that were being developed at the time of cancellation, and is planning to try again. Weisskopf and Costa, meanwhile, are planning to submit a mission concept called XIPE (X-ray Imaging Polarimetry Explorer), which was passed over in 2012 by the European Space Agency.

Either mission could measure the spin of a black hole, which imprints spin geometry on the polarized X-ray light it emits as swirling disks of matter fall into it. The polarimeters could also be used to settle a debate over two theories that describe where and how X-rays are emitted in the atmosphere of pulsars.

NASA astrophysics director Paul Hertz says that the agency remains open to an X-ray polarimetry mission despite cancelling GEMS last year. The competition’s winning mission, which would launch by the end of 2020, will depend on peer review, he says. Dozens of proposals are expected from many fields of astrophysics, including exoplanet research — an extremely hot area. But Weisskopf is optimistic that at last the field he founded in the 1970s can spring to life after being neglected for so long. “The time has come,” he says. ■

CORRECTIONS

In the News Feature ‘BRAIN storm’ (*Nature* **503**, 26–28; 2013), a description that characterized a May meeting run by the National Science Foundation as being chaotic was incorrectly attributed to Van Wedeen. And in the News Feature ‘A race against resistance’ (*Nature* **503**, 186–188; 2013), the quote “If we just roll this out without surveillance, we risk repeating all of the mistakes made in the past” was wrongly attributed to Paul Milligan instead of to Christopher Plowe.



*Graphene's dazzling properties
promise a technological
revolution, but it may take a
billion euros to overcome some
fundamental problems.*

BY MARK PELOW

Mr G gazes out from a recruitment poster hanging in an engineering building in Cambridge, UK. His cartoon cape billows out behind him, his sketched-in muscles ripple beneath his costume, his chest is emblazoned with a 'G' inside a hexagon — and his forefinger points straight at the viewer. “I want you for the Graphene Flagship!” declares the cartoon crusader, championing a material as super as he is.

Graphene is the thinnest substance ever made: a single sheet of carbon atoms arranged in a hexagonal honeycomb pattern. It is as stiff as diamond and hundreds of times stronger than steel — yet at the same time is extremely flexible, even stretchable. It conducts electricity faster at room temperature than any other known material, and it can convert light of any wavelength into a current. In the decade since graphene was first isolated, researchers have proposed dozens of potential applications, from faster computer chips and flexible touchscreens to hyper-efficient solar cells and desalination membranes.

But harnessing graphene's qualities for practical use has proved a massive challenge. Graphene is complicated and expensive to make in large sheets, which usually have so many atomic-scale flaws and tears that they fail to match the amazing properties of the tiny flakes studied in the laboratory. And even if its quality were good, there are no well-established industrial methods for handling something so thin, or for integrating it with other materials to create useful products. What's more, graphene has

a superweakness. Its electrons may be extremely mobile, but other properties make it fundamentally unsuitable for the sort of on–off switching that lies at the heart of digital electronics.

Hence Mr G's call to arms. The character was created in 2011 to help publicize a multinational push for a Graphene Flagship project: a decade-long, €1-billion (US\$1.35-billion), all-European effort to take graphene from the laboratory bench to the factory floor. And not just graphene. The project's proponents also wanted to study more than a dozen other atomically thick materials discovered in graphene's wake — that, when sandwiched together with graphene, might help to overcome its limitations¹.

The campaign worked: the European Commission in Brussels gave its go-ahead to the graphene flagship project in January (see *Nature* **493**, 585–586; 2013). Already the world's largest research effort on the material, encompassing hundreds of scientists across 17 European countries, it will grow even larger after the flagship puts out its first call for additional project proposals on 25 November.

The infusion of funds and energy has galvanized the graphene community, says Andrea Ferrari, director of the Cambridge Graphene Centre and chair of the flagship's executive board. Ferrari, whose office wall sports Mr G's poster, says “Nobody has been involved in anything this big before,”

TOO MANY COOKS?

But some question whether the programme is too big. Is an academia–industry collaboration, inevitably fettered by the bureaucracy of such a large venture, the best way to deliver a technological revolution? “This is not the way products are actually developed,” says Phaedon Avouris, a graphene and nanotechnology researcher at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York. And some researchers involved in the project are concerned that political forces, rather than scientific priorities, will steer the dispersal of funds over the next few years.

Still, the flagship's prospects for success seem strong enough that national governments and industry partners, such as Nokia and Airbus, will collectively put up half its funding. (The European Commission will provide the rest.) “I hope that after ten years, technologies based on graphene or other layered materials are mainstream,” says the flagship's director Jari Kinaret, who is based at Chalmers University of Technology in Gothenburg, Sweden. Just as we now do with polymers, semiconductors and ceramics, he says, “we should take graphene for granted”.

The flagship programme is divided into 16 work packages, most of them targeted at developing applications such as high-frequency

electronics, sensors and energy storage. Next week's call for proposals, worth €9 million, comes at the beginning of a €54-million ramp-up phase that is expected to deliver the first wave of prototypes by 2016.

But there will be no graphene computer chips, graphene sensors or graphene solar cells without a steady supply of graphene itself. One of the flagship's first and biggest challenges is to find more economical and reliable ways to produce high-quality sheets of the material.

Most research laboratories still make graphene using the method pioneered in 2004 by Andre Geim and Konstantin Novoselov at the University of Manchester, UK, who went on to win the 2010 Nobel Prize in Physics for their studies. Geim and Novoselov found that they just had to touch a strip of household sticky tape to ordinary graphite — which consists of billions of layers of graphene stacked on top of one another — and they could peel off thin flakes of carbon. By repeatedly splitting those flakes, they were eventually left with graphene². This was a technique that any laboratory could use, and graphene research exploded.

But the method is much too slow and finicky for industrial-scale production. Just one micrometre-sized flake made in this way can cost more than \$1,000 — making it, gram for gram, one of the most expensive materials on Earth.

The leading alternative³ relies on chemical vapour deposition (CVD), whereby methane is piped over a catalytic copper foil heated to about 1,000 °C. As the methane breaks down, small islands of pure carbon begin to grow on the foil, linking together to form a patchwork polycrystalline sheet of graphene. Harsh chemicals are then used to etch away the copper to free a sheet of graphene tens of centimetres wide, which can be transferred to a silica or polymer substrate. That process brings costs below \$100,000 per square metre, but the product is often riddled with defects, impairing its electrical properties and making it much weaker than flakes produced by the sticky-tape method.

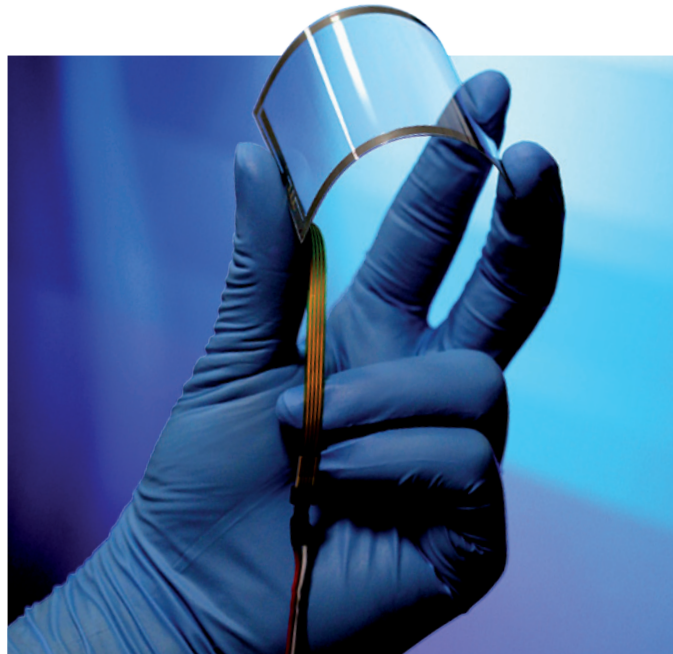
INDUSTRIAL ACTION

The flagship programme is tackling this problem in part through its industrial partners, such as Graphenea of San Sebastián, Spain, which already makes about 15 square metres of graphene per year. And it should benefit from a deal signed in September that will see fledgling graphene producer Bluestone Global Tech of Wappingers Falls, New York, open a pre-production facility and offices at the National Graphene Institute in Manchester, the hub of Britain's graphene effort. This year, Bluestone began speeding up production and lowering costs by using bubbles of hydrogen to tease large graphene monolayers away from the copper foil without etching^{4,5}.

Yet even Bluestone's manufacturing process is "still a very complex way of adding graphene to a substrate", says Tapani Ryhänen, head of sensors and material research for Finnish company Nokia, and a member of the flagship's advisory council. The flagship aims to refine the CVD process and to improve on alternative production methods. Also problematic is the tricky process of transferring the freshly made graphene from its catalytic foil to a new substrate. Lay it on top of silicon, for example, and the sheet wrinkles and puckers. One solution would be to grow graphene directly on the substrate, or on top of another sturdy, protective monolayer such as boron nitride, a process demonstrated at small scale earlier this year⁶.

But ultimately, says Rod Ruoff at the University of Texas at Austin who led the development of the CVD production method, the best way to slash costs and propel graphene into the mainstream would be to make high-quality monolayers from bulk graphite — exfoliation on an industrial scale. The flagship will investigate chemical treatments, ultrasonic vibration and more, but a practical, scalable method still seems a long way off. "We need some sort of a breakthrough here," says Ruoff.

Despite its manufacturing challenges, enthusiasts are quick to point out that graphene has already hit the market. Multi-layered graphene, in which many sheets are stacked together, is used to strengthen a tennis



BYUNG HEE HONG

Graphene offers a way to make flexible and transparent smartphone screens.

racquet made by Head, for example, and forms a conductive circuit in anti-theft packaging produced by Vorbeck Materials in Jessup, Maryland.

But these cheaper forms of graphene include a range of different structures that are essentially nanometre-sized chunks of graphite. The properties of this sooty jumble of fragments are no match for Mr G's superpowers, which reach their zenith only in pristine, one-atom-thick layers in which the atomic arrangement is perfect. Only in this state can electrons flow more quickly than in any other material.

To get current moving through any crystal, electrons must first clear a hurdle called the band gap: the energy required to knock them loose from individual atoms and set them free to roam. Insulating materials have a large band gap, meaning that electrons tend to be tightly bound to the atoms and need a huge kick to start moving (see 'Mind the gap'). Semiconductors such as silicon and germanium have a much smaller band gap, so only a little jolt of energy is required. Metals have no band gap at all; they are great conductors because at least some of their electrons are always free. But graphene sits right on the boundary, blessed with an infinitesimally small band gap that helps current to zip across its interlocking hexagons 100 to 200 times faster than it can move through silicon⁷.

This tiny band gap also makes graphene optically omnivorous. Silicon can only absorb photons with energies greater than its band gap; if weaker photons hit it, they can't free electrons from the parent atoms. Graphene, by contrast, can absorb photons across the visible spectrum and beyond, turning their energy into electrical current. "There's not really another material that has good properties for both optics and electronics," says Daniel Neumaier of the contract research company AMO in Aachen, Germany, who is leading the flagship's high-frequency electronics work package.

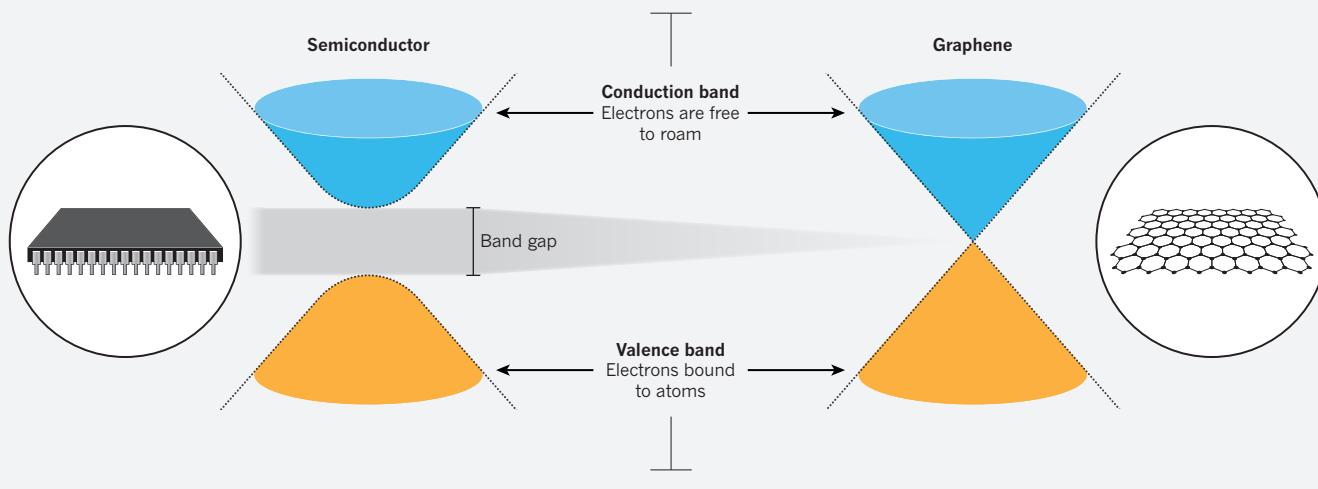
This combination of abilities makes graphene a promising candidate for converting photons into electrical signals. Graphene photodetectors could allow computer chips to communicate with light rather than comparatively sluggish, energy-wasting electrons — an advance that would cut power consumption and allow computers to handle data more efficiently. Such photodetectors would be smaller than current devices made of germanium, and could handle a wide range of wavelengths, allowing them to interpret multiple signals bundled together into the same beam (see *Nature* <http://doi.org/pz2>; 2013).

Graphene could also be useful in medical and security scanning that

"Nobody will ditch silicon unless there's a really compelling reason to do so."

MIND THE GAP

Electrons in a solid are restricted to certain ranges, or bands, of energy (vertical axis). In an insulator or semiconductor, an electron bound to an atom can break free only if it gets enough energy from heat or a passing photon to jump the 'band gap', but in graphene the gap is infinitesimal. This is the main reason why graphene's electrons can move very easily and very fast.



uses terahertz-frequency radiation. The generation and manipulation of terahertz waves, which lie between the infrared and microwave regions of the spectrum, often requires bulky equipment or cryogenic cooling. Graphene-based devices are compact and can generate or detect the waves at room temperature. This may be graphene's best opportunity for a groundbreaking application, says Avouris, because it could find a role not already occupied by other well-established materials.

Others think that graphene's most obvious optical property — its transparency — may yield its first major application in the electronics industry. Samsung and other Asian companies are developing transparent graphene electrodes to serve as smartphone touchscreens. The indium tin oxide electrodes in use today are brittle, whereas graphene is strong and flexible. And although graphene touchscreens are currently more expensive than the conventional variety, "the cost is falling rapidly as we ramp up the scale of production", says Bluestone's co-founder Yu-Ming Lin.

TURN OFF

When it comes to digital electronics, however, graphene's greatest strength is also its greatest weakness. In principle, its extremely mobile electrons could allow graphene transistors to process data at very high rates, with some devices already clocking in at more than 400 gigahertz — many times faster than comparable silicon devices⁸. But graphene's lack of band gap makes it very hard to turn the current off once it starts flowing, a serious impediment to logic operations, which are all about on-off switching. Doping graphene with other materials or slicing it into narrow ribbons can open up a small band gap, but this also slows the flow of electrons. So researchers are trying to tune its electrical properties by combining graphene with other monolayer materials such as boron nitride or creating transistors from molybdenum disulphide and tungsten diselenide^{9–11}.

But graphene is still a long way from replacing silicon electronics, says Tim Harper of the technology-development company Cientifica, based in London: "Nobody will just ditch silicon unless there's a really compelling reason to do so." In the near term, a graphene transistor's biggest selling point may be its ability to operate over a range of voltages, rather than any ability to switch on or off. Applications might include sensors for environmental pollutants or blood-oxygen levels, or the transmitters and receivers inside mobile phones. By the end of the programme's 30-month ramp-up phase, Neumaier's goal is to build prototypes that demonstrate graphene's potential in these areas. "Expectations at the moment are very large," he says.

So are the concerns of some researchers. As one

of Europe's highest-profile science projects, the graphene flagship has some treacherous political waters to navigate.

The European Commission wants the flagship to be as inclusive as possible, to ensure that under-represented member states get a piece of the action. One consequence is that next week's call is open only to new partners — existing flagship research groups are barred from bidding for those funds. "That came as a surprise," says Kinaret. The rule excludes all researchers who have signed up en masse through national research networks, including the CNRS in France, the Max Planck Society in Germany and the CSIC in Spain. The networks have lobbied the commission to change that rule, but "we have been less than successful", says Kinaret.

Kinaret expects that restriction to change next year after the European Union's Horizon 2020 research programme comes into force, and other funding streams are available in the meantime. But some researchers have been left with a sense of foreboding. Ferrari worries that there is a risk of losing sight of the original goal: a genuine technological revolution in ten years. By slicing the flagship's money into smaller chunks and spreading it more widely, Europe could keep more member states happy — but might dilute the project's impact. "Excellence must be the criterion," he insists.

Meanwhile, Europe faces stiff competition from Asia in the race to commercialize graphene. Although the European Union leads the world in academic publications on the material, the UK government's Intellectual Property Office in Newport reported in March that 15 of the top 20 global graphene patent-holders are Chinese, Japanese and South Korean companies and universities, with Samsung way out in front. Some Chinese manufacturers say that mobile devices bearing graphene touchscreens will hit the market next year.

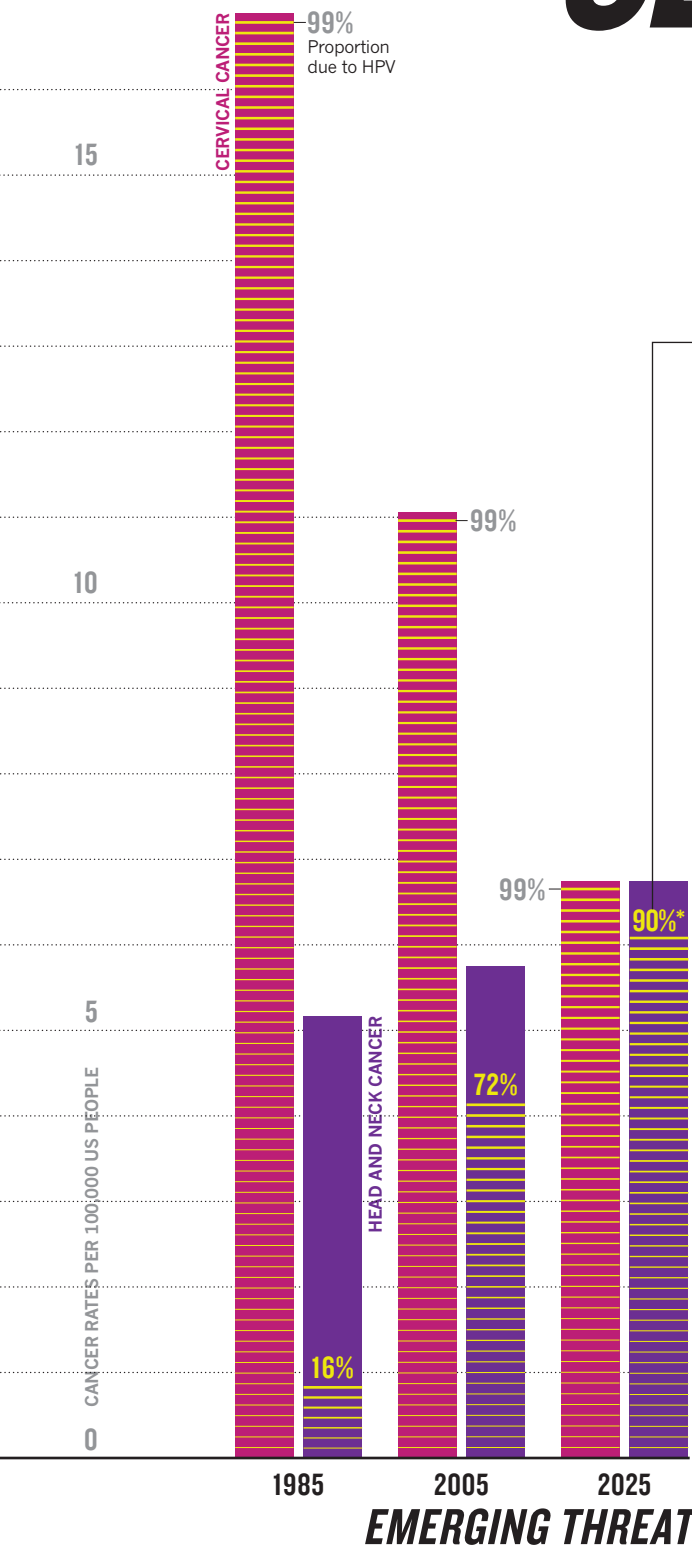
Europe has led in academic research on graphene, but it trails in development. "That," says Kinaret, "is what we are hoping to change." ■

Mark Peplow is a freelance writer based in Cambridge, UK.

1. Geim, A. K. & Grigorieva, I. V. *Nature* **499**, 419–425 (2013).
2. Novoselov, K. S. *et al. Science* **306**, 666–669 (2004).
3. Li, X. *et al. Science* **324**, 1312–1314 (2009).
4. Gao, L. *et al. Nature Commun.* **3**, 699 (2012).
5. Wang, Y. *et al. ACS Nano* **5**, 9927–9933 (2011).
6. Yang, W. *Nature Mater.* **12**, 792–797 (2013).
7. Chen, J.-H., Jang, C., Xiao, S., Ishigami, M. & Fuhrer, M. S. *Nature Nanotechnol.* **3**, 206–209 (2008).
8. Cheng, R. *et al. Proc. Natl Acad. Sci. USA* **109**, 11588–11592 (2012).
9. Hunt, B. *et al. Science* **340**, 1427–1430 (2013).
10. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. *Nature Nanotechnol.* **6**, 147–150 (2011).
11. Liu, W. *et al. Nano Lett.* **13**, 1983–1990 (2013).

➔ **NATURE.COM**
For more on
graphene, see
Nature's Outlook:
go.nature.com/hm41sm

SEX, CANCER and a VIRUS



Rates of head and neck cancer (purple) have risen — and they are set to grow further. An increasing proportion of cases is caused by human papillomavirus (HPV, yellow). At the same time, rates of cervical cancer (red; nearly all caused by HPV) have declined, owing to increased screening.

*Estimate based on clinical observations

Human papillomavirus is causing a new form of head and neck cancer — leaving researchers scrambling to understand risk factors, tests and treatments.

BY MEGAN SCUDELLARI

On a sunny day in 1998, Maura Gillison was walking across the campus of Johns Hopkins University in Baltimore, Maryland, thinking about a virus. The young oncologist bumped into the director of the university's cancer centre, who asked politely about her work. Gillison described her discovery of early evidence that human papillomavirus (HPV) — a ubiquitous pathogen that infects nearly every human at some point in their lives — could be causing tens of thousands of cases of throat cancer each year in the United States. The senior doctor stared down at Gillison, not saying a word. "That was the first clue that what I was doing was interesting to others and had potential significance," recalls Gillison.

She knew that such a claim had a high burden of proof. HPV was known to cause cervical cancer and small numbers of genital cancers, but no other forms. So Gillison started a careful population study comparing people with cancer to healthy individuals. Over seven years, she recruited 300 participants, collected tissue samples, and never once looked at the data. "My policy, when doing a study, is that we wait until all the data are in, and do all the analyses at once," says Gillison, who is as careful as she is blunt. "I don't know anything until the data tell me."

Only in 2005 did Gillison finally sit down with a doctoral student to analyse the data. Within an hour, the fruits of those years of labour popped up on the computer screen: people with head and neck cancer were 15 times more likely to be infected with HPV in their mouths or throats than those without¹. The association backed up some of Gillison's earlier work, which showed² how HPV DNA integrates itself into the nuclei of throat cells and produces cancer-causing proteins. Gillison leapt from her chair and began jumping up and down. "The association was so incredibly strong, it made me realize this was absolutely irrefutable evidence," she says.

Since then, she and a network of other researchers have amassed a mountain of evidence that HPV causes a large proportion of head and neck cancers, and that these HPV-positive cancers are on the rise. The finding has been "a paradigm-shifting realization in the field", says Robert Ferris, chief of the division of head and neck surgery at the University of Pittsburgh Cancer Institute in Pennsylvania.

The medical community is struggling to come to grips with the implications. There is currently no good screening method for HPV-caused

SOURCE: M. GILLISON/A. K. CHATURVEDI ET AL. J. CLIN. ONCOL. 29, 4294-4301 (2011)

cancer in the head and neck, and commercially available HPV vaccines are still prescribed only to people under the age of 26, despite evidence that they could prevent head and neck cancer in all adults. Plus, if HPV can get into the mucous membranes of the mouth and throat, where does it stop? There are hints that HPV is a risk factor for other, even more common, types of cancer, including lung cancer.

For now, researchers and doctors need to learn more about how HPV causes cancer, and how best to prevent and treat it, says Gillison. "Our clinics are flooded" with head and neck cancers triggered by HPV, she says, vexation clear in her voice. "But though I talk about it constantly in public settings and the lay press, it amazes me that it's often as if no one has heard of it."

NEW THREAT

James Rocco, director of head and neck molecular oncology research at Massachusetts General Hospital in Boston, remembers the first signs that something was changing. Until the late 1990s, most cases of cancer in the back of the throat (the oropharynx) could be blamed on alcohol and tobacco use: the majority of Rocco's patients were men around 50 years old, who had been smoking and drinking for 30 years. But then 40-year-old marathon runners and people in otherwise good health began to trickle — then stream — into his office. And when treated with chemotherapy and radiation, these people seemed to have better survival rates than the other head and neck cancer patients.

There were also irregularities in the laboratory. When biopsied, the site of the cancer was slightly different in this healthier cohort: instead of beginning on the surface of the tonsil as normal, tumours seemed to start deep in tonsil crevices. And more and more of the tumours lacked mutations in a protein called p53 — then considered a hallmark of oropharyngeal cancer. "We kind of knew we were dealing with something different," recalls Rocco.

Gillison started pursuing the issue in 1996, after a passing comment by a colleague. Keerti Shah, a molecular microbiologist at the Johns Hopkins Bloomberg School of Public Health, had mentioned research in Finland that had identified HPV in a cell line developed from an oropharyngeal tumour³. As Shah and Gillison walked around campus one day, they talked about the finding. Was it an isolated case? Had HPV contaminated the sample? Or, as Shah suspected, could HPV cause some cases of head and neck cancer?

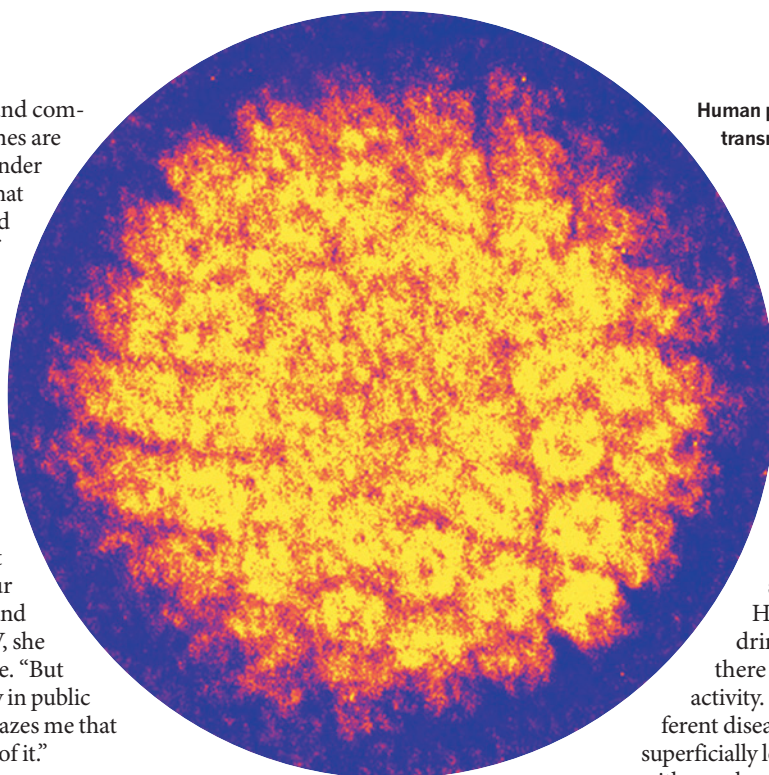
Gillison went straight to her office to do a literature search. She began analysing tumour samples from the Head and Neck Cancer

If HPV can get into the mucous membranes of the mouth and throat, where does it stop?

Center at Hopkins and found HPV in about 25% of them. She used multiple techniques to be sure that positive results were not attributable to laboratory contamination. She looked for the virus in early, middle and late stage tumours. HPV was not just present; she found that its DNA had infiltrated the tumours and was producing two potent oncoproteins, an indication it was the cause of the cancer. Gillison also profiled people with HPV to learn about the cancer's clinical characteristics, and identified molecular biomarkers that were absent in tumours without HPV. She worked on the project for 18 months, without taking a day off.

She, Shah and their colleagues published their results in 2000 (ref. 2), demonstrating that HPV-positive oropharyngeal cancer is a distinct type of cancer that starts deep in the tonsils, has HPV DNA present in the tumour-cell nuclei but not neighbouring cells, has fewer p53 mutations than HPV-negative cancer, has less association with smoking and alcohol consumption and has better survival rates. But many oncologists were sceptical: some suspected that HPV was just a passenger virus, or that its presence was the result of contamination. Others thought that HPV might be just a risk factor, rather than a cause, for head and neck cancer — one of several ingredients, including drinking and smoking, that when combined together congealed into a cancerous stew.

Human papillomavirus, seen in a coloured transmission electron micrograph.



In 2007, Gillison published her seven-year population study showing the link between oral HPV infection and oropharyngeal cancer¹; the next year, she released a study⁴ showing that HPV-positive and HPV-negative oropharyngeal cancers had completely different risk profiles. People with HPV-positive cancer tended to have had many oral-sex partners, but there was no statistical association with tobacco smoking or drinking; those with HPV-negative cancers were heavy drinkers and cigarette smokers but there was no association with sexual activity. "These were two completely different diseases," says Gillison. "They might superficially look similar — a patient comes in with a neck mass and their throat hurts — but I realized what drove the pathogenesis was completely different in the two cases."

By then, all doubts had faded. In 2007, the World Health Organization's International Agency for Research on Cancer in Lyons, France, declared that there was sufficient evidence to conclude that HPV causes a subset of oropharyngeal cancers. Gillison's research has been "definitive", says Jeffrey Myers, director of head and neck surgery research at the University of Texas MD Anderson Cancer Center in Houston.

Community acceptance came not a moment too soon. The number of oropharyngeal cancers has been growing over the past 30 years: there are now 10,000 cases in the United States each year, a number that is likely to climb to 16,000 by 2030 (see 'Emerging threat'). An overwhelming majority are caused by HPV. Worldwide, cancer centres report that the virus is responsible for between 45% and 90% of oropharyngeal cancers. "In Europe, HPV-positive oropharyngeal cancers have almost quadrupled in number over a period of 10 to 15 years," says Hisham Mehanna, director of the Institute of Head and Neck Studies and Education at the University of Birmingham, UK, who has published a meta-analysis⁵ of more than 250 papers on prevalence rates. "Our projection suggests that it's going to continue to increase significantly." Why rates are escalating is unknown, although one suggestion points to increasing numbers of sexual partners.

PROBLEM PROTEINS

It turns out that HPV causes throat cancer in much the same ways as it causes cancer in the cervix. The virus's DNA integrates into human DNA in the nuclei of healthy cells, and uses the cells' machinery to produce two harmful proteins, E6 and E7. These bind to, and shut down,

two important tumour-suppressor proteins, p53 and pRb. Active pRb prevents excessive cell growth; without it, cells proliferate unchecked. Active p53 arrests the cell-division cycle when DNA is damaged, and then either activates DNA repair or initiates cell death. Without p53, a cell replicates wildly even if it has DNA damage.

In cancers caused by HPV, the virus silences p53 but leaves the gene that produces it intact; by contrast, in HPV-negative cancers, the gene is mutated, probably through exposure to carcinogens, and produces an ineffective version of the protein. This may explain why people with HPV-positive oropharyngeal cancer respond better to treatment: early evidence suggests⁶ that chemotherapy or radiation may somehow reactivate p53 in HPV-positive cancers, turning the powerful protein back on to fight the tumour.

There are other possibilities. It could be that people with HPV-positive cancer are generally healthier than their HPV-negative counterparts: they tend to be younger, generally don't smoke and are more likely to comply with treatment regimes. Another possibility, supported by a study⁷ using sequencing data from 74 head and neck cancers, is that HPV-negative tumours are more heterogeneous than HPV-positive tumours. The cells have many more mutations, and a wider range of them. In an HPV-negative tumour, therefore, "there's more likely to be something in there that will resist therapy", says Rocco, a co-author of the study.

TOXIC TREATMENT

The fact that people with HPV-positive cancer have better outcomes has caused many clinicians, including Gillison and Ferris, to wonder whether these patients should get different treatments. The current standard therapy for oropharyngeal cancer is a combination of cisplatin — a toxic, potent chemotherapy drug — and radiation. This has many potential side effects, including damage to the voice box and throat, which can hinder the ability to speak and swallow. With the younger, healthier HPV-positive patients, who are 58% less likely to die within three years of treatment than HPV-negative patients, clinicians worry about the long-term effects of the treatment, and are exploring techniques including less-toxic chemotherapy regimens.

Researchers are also looking at ways to prevent the disease in the first place. More than 90% of HPV-related oropharyngeal cancers are caused by HPV-16, a particularly dangerous strain and the main cause of cervical cancer. The two vaccines approved to prevent cervical cancer, Merck's Gardasil and GlaxoSmithKline's Cervarix, both protect against HPV-16. In theory, therefore, protection against HPV-positive oropharyngeal cancer is already in doctors'



"These diseases might look similar, but what drove the pathogenesis was completely different."

Maura Gillison

cabinets. A clinical trial of 5,840 women, published this year by researchers at the US National Cancer Institute⁸, showed that Cervarix is 93% effective at preventing oral HPV infection in both women with pre-existing cervical infections and those without, none of whom had been previously vaccinated.

A major barrier stands in the way of official approval for using the vaccine to protect against oropharyngeal cancer: there is not yet a way to prove that it would work. For cervical cancer, doctors test cells taken from the cervix during routine screening, looking for changes that precede the emergence of cancer. Because HPV-positive oropharyngeal cancer arises deep in the tonsil, checks would have to be much more invasive. "In theory, we could detect it, but we would need to do a tonsillectomy on everyone in the vaccine trial," says Gillison. "That's never going to happen."

There may be another way. Mehanna and his colleagues are in the process of analysing the tonsils of 1,250 people who underwent tonsillectomies for non-cancerous reasons. The researchers have identified what they think are pre-malignant lesions in some HPV-positive samples that may represent the earliest stages of the cancer, and could serve as a biomarker. "We're now testing to make sure this pre-malignancy is driven by HPV and is not just random," says Mehanna.

Other concerns and questions linger. For

example, scientists have yet to determine whether oral HPV infection comes only from sexual acts that involve contact between the mouth and genitals, or also from other acts including deep kissing. And most people who develop an HPV infection do not get oropharyngeal cancer: about 90% of those who become infected orally clear the infection within two years. No one is sure why.

Researchers are also investigating whether HPV causes other types of cancer. There have been studies of the relationship between the virus and oesophageal cancer, but findings have been inconclusive. Another area of interest is the lung.

There, too, tobacco has been the primary culprit for decades, but some 15–20% of lung-cancer cases in men and 50% in women are in people who have never smoked. Doctors have theorized that a virus lies behind them.

The available data are conflicting. One paper⁹ in 2001 identified HPV DNA in 55% of 141 lung tumours, compared with 27% of 60 non-cancer control samples. And in 2009, researchers led by Iver Petersen, director of the Institute for Pathology at Jena University Hospital in Germany, conducted a meta-analysis¹⁰ of 53 publications examining 4,508 cases of lung cancer, and concluded that "HPV is the second most important cause of lung cancer after cigarette smoking". They encouraged more research. But many other studies have refuted those observations, including one from Gillison and her colleagues, in which they used sensitive DNA assays to study the lung cancers of 450 patients, and found no HPV (ref. 11).

With head and neck cancer, however, Gillison is optimistic that new knowledge about HPV as a cause of the disease will help physicians to treat it — and eventually to prevent it with a vaccine. "In terms of cancer," she says, "there aren't many populations where we've identified the necessary cause and have a potential solution on the shelf." ■

Megan Scudellari is a freelance reporter based in Boston, Massachusetts.

1. D'Souza, G. *et al.* *New Engl. J. Med.* **356**, 1944–1956 (2007).
2. Gillison, M. L. *et al.* *J. Natl Cancer Inst.* **92**, 709–720 (2000).
3. Syrjänen, K. J., Pyrhönen, S., Syrjänen, S. M. & Lamberg, M. A. *Br. J. Oral Surg.* **21**, 147–153 (1983).
4. Gillison, M. L. *et al.* *J. Natl Cancer Inst.* **100**, 407–420 (2008).
5. Mehanna, H. *et al.* *Head Neck* **35**, 747–755 (2013).
6. Xie, X. *et al.* *Oncogene* <http://dx.doi.org/10.1038/ncr.2013.25> (2013).
7. Mroz, E. A. & Rocco, J. W. *Oral Oncol.* **49**, 211–215 (2013).
8. Herrero, R. *et al.* *PLoS ONE* **8**, e68329 (2013).
9. Cheng, Y. W. *et al.* *Cancer Res.* **61**, 2799 (2001).
10. Klein, F., Amin Kotb, W. F. M. & Petersen, I. *Lung Cancer* **65**, 13–18 (2009).
11. Koshiol, J. *et al.* *J. Natl Cancer Inst.* **103**, 501–507 (2011).

COMMENT

POLICY Twenty tips to help non-scientists to assess evidence **p.335**



FICTION *Brave New World* reappraised 50 years after Aldous Huxley's death **p.338**

HISTORY How science and technology have changed surfing **p.341**

EDUCATION Data show massive open online courses mainly reach educated males **p.342**

PAUL BLOW



The risks of the replication drive

The push to replicate findings could shelve promising research and unfairly damage the reputations of careful, meticulous scientists, says **Mina Bissell**.

Every once in a while, one of my postdocs or students asks, in a grave voice, to speak to me privately. With terror in their eyes, they tell me that they have been unable to replicate one of my laboratory's previous experiments, no matter how hard they try. Replication is always a concern when dealing with systems as complex as the three-dimensional cell cultures routinely used in my lab. But with time and careful consideration of experimental conditions, they, and others, have always

managed to replicate our previous data.

Articles in both the scientific and popular press^{1–3} have addressed how frequently biologists are unable to repeat each other's experiments, even when using the same materials and methods. But I am concerned about the latest drive by some in biology to have results replicated by an independent, self-appointed entity that will charge for the service. The US National Institutes of Health is considering making validation routine for certain types of experiments, including the basic science that

leads to clinical trials⁴. But who will evaluate the evaluators? The Reproducibility Initiative, for example, launched by the journal *PLoS ONE* with three other companies, asks scientists to submit their papers for replication by third parties, for a fee, with the results appearing in *PLoS ONE*. *Nature* has targeted⁵ reproducibility by giving more space to methods sections and encouraging more transparency from authors, and has composed a checklist of necessary technical and statistical information. This should be applauded. ►

► So why am I concerned? Isn't reproducibility the bedrock of the scientific process? Yes, up to a point. But it is sometimes much easier not to replicate than to replicate studies, because the techniques and reagents are sophisticated, time-consuming and difficult to master. In the past ten years, every paper published on which I have been senior author has taken between four and six years to complete, and at times much longer. People in my lab often need months — if not a year — to replicate some of the experiments we have done on the roles of the microenvironment and extracellular matrix in cancer, and that includes consulting with other lab members, as well as the original authors.

People trying to repeat others' research often do not have the time, funding or resources to gain the same expertise with the experimental protocol as the original authors, who were perhaps operating under a multi-year federal grant and aiming for a high-profile publication. If a researcher spends six months, say, trying to replicate such work and reports that it is irreproducible, that can deter other scientists from pursuing a promising line of research, jeopardize the original scientists' chances of obtaining funding to continue it themselves, and potentially damage their reputations.

FAIR WIND

Twenty years ago, a reproducibility movement would have been of less concern. Biologists were using relatively simple tools and materials, such as pre-made media and embryonic fibroblasts from chickens and mice. The techniques available were inexpensive and easy to learn, thus most experiments would have been fairly easy to double-check. But today, biologists use large data sets, engineered animals and complex culture models, especially for human cells, for which engineering new species is not an option.

Many scientists use epithelial cell lines that are exquisitely sensitive. The slightest shift in their microenvironment can alter the results — something a newcomer might not spot. It is common for even a seasoned scientist to struggle with cell lines and culture

conditions, and unknowingly introduce changes that will make it seem that a study cannot be reproduced. Cells in culture are often immortal because they rapidly acquire epigenetic and genetic changes. As such cells divide, any alteration in the media or microenvironment — even if minuscule — can trigger further changes that skew results. Here are three examples from my own experience.

My collaborator, Ole Petersen, a breast-cancer researcher at the University of Copenhagen, and I have spent much of our scientific careers learning how to maintain the functional differentiation of human and mouse mammary epithelial cells in culture. We have succeeded in cultivating human breast cell lines for more than 20 years, and when we use them in the three-dimensional assays that we developed^{6,7}, we do not observe functional drift. But our colleagues at biotech company Genentech in South San Francisco, California, brought to our attention that they could not reproduce the architecture of our cell colonies, and the same cells seemed to have drifted functionally. The collaborators had worked with us in my lab and knew the assays intimately. When we exchanged cells and gels, we saw that the problem was in the cells, procured from an external cell bank, and not the assays.

Another example arose when we submitted what we believe to be an exciting paper for publication on the role of glucose uptake in cancer progression. The reviewers objected to many of our conclusions and results because the published literature strongly predicted the prominence of other molecules and pathways in metabolic signalling. We then had to do many extra experiments to convince them that changes in media glucose levels, or whether the cells were in different contexts (shapes) when media were kept constant, drastically changed the nature of the metabolites produced and the pathways used⁸.

A third example comes from a non-malignant human breast cell line that is now used by many for three-dimensional experiments. A collaborator noticed that her group could not reproduce its own data convincingly when using cells from a cell bank. She had obtained the original cells from another investigator. And they had been cultured

under conditions in which they had drifted. Rather than despairing, the group analysed the reasons behind the differences and identified crucial changes in cell-cycle regulation in the drifted cells. This finding led to an exciting, new interpretation of the data that were subsequently published⁹.

REPEAT AFTER ME

The right thing to do as a replicator of someone else's findings is to consult the original authors thoughtfully. If e-mails and phone calls don't solve the problems in replication, ask either to go to the original lab to reproduce the data together, or invite someone from their lab to come to yours. Of course replicators must pay for all this, but it is a small price in relation to the time one will save, or the suffering one might otherwise cause by declaring a finding irreproducible.

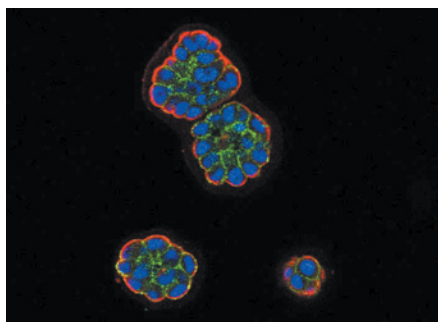
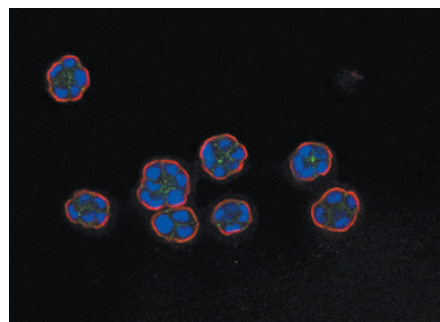
When researchers at Amgen, a pharmaceutical company in Thousand Oaks, California, failed to replicate many important studies in preclinical cancer research, they tried to contact the authors and exchange materials. They could confirm only 11% of the papers³. I think that if more biotech companies had the patience to send someone to the original labs, perhaps the percentage of reproducibility would be much higher.

It is true that, in some cases, no matter how meticulous one is, some papers do not hold up. But if the steps above are taken and the research still cannot be reproduced, then these non-valid findings will eventually be weeded out naturally when other careful scientists repeatedly fail to reproduce them. But sooner or later, the paper should be withdrawn from the literature by its authors.

One last point: all journals should set aside a small space to publish short, peer-reviewed reports from groups that get together to collaboratively solve reproducibility problems, describing their trials and tribulations in detail. I suggest that we call this ISPA: the Initiative to Solve Problems Amicably. ■

Mina Bissell is Distinguished Scientist in the Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.
e-mail: mjbissell@lbl.gov

JAMIE INMAN/BISSELL LAB



Cells from the same human breast cell line from different sources respond differently to the same assay.

1. Naik, G. 'Scientists' Elusive Goal: Reproducing Study Results' *The Wall Street Journal* (2 December 2011); available at <http://go.nature.com/aqopcs>.
2. *Nature Med.* **18**, 1443 (2012).
3. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012).
4. Wadman, M. *Nature* **500**, 14–16 (2013).
5. *Nature* **496**, 398 (2013).
6. Barcellos-Hoff, M. H., Aggeler, J., Ram, T. G. & Bissell, M. J. *Development* **105**, 223–235 (1989).
7. Petersen, O. W., Rønnov-Jessen, L., Howlett, A. R. & Bissell, M. J. *Proc. Natl Acad. Sci. USA* **89**, 9064–9068 (1992).
8. Onodera, Y., Nam, J.-M. & Bissell, M. J. *J. Clin. Invest.* (in the press).
9. Ordinario, E. et al. *PLoS ONE* **7**, e51786 (2012).



Twenty tips for interpreting scientific claims

This list will help non-scientists to interrogate advisers and to grasp the limitations of evidence, say **William J. Sutherland, David Spiegelhalter and Mark A. Burgman.**

Calls for the closer integration of science in political decision-making have been commonplace for decades. However, there are serious problems in the application of science to policy — from energy to health and environment to education.

One suggestion to improve matters is to encourage more scientists to get involved in politics. Although laudable, it is unrealistic to expect substantially increased political involvement from scientists. Another proposal is to expand the role of chief scientific advisers¹, increasing their number, availability and participation in political processes. Neither approach deals with the core problem of scientific ignorance among many who vote in parliaments.

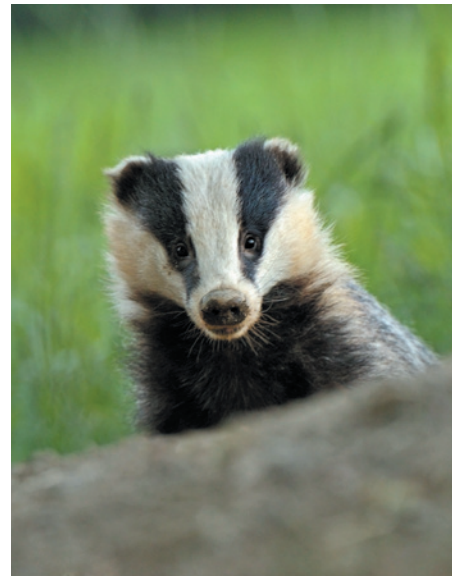
Perhaps we could teach science to politicians? It is an attractive idea, but which busy politician has sufficient time? In practice, policy-makers almost never read scientific papers or books. The research relevant to the topic of the day — for example, mitochondrial replacement, bovine tuberculosis or nuclear-waste disposal — is interpreted for them by advisers or external advocates. And there is rarely, if ever, a beautifully designed double-blind, randomized, replicated, controlled experiment with a large sample size and unambiguous conclusion that tackles the exact policy issue.

In this context, we suggest that the immediate priority is to improve policy-makers' understanding of the imperfect nature of science. The essential skills are to be able to intelligently interrogate experts and advisers, and to understand the quality, limitations and biases of evidence. We term these interpretive scientific skills. These skills are more accessible than those required to understand the fundamental science itself, and can form part of the broad skill set of most politicians.

To this end, we suggest 20 concepts that should be part of the education of civil servants, politicians, policy advisers and journalists — and anyone else who may have to interact with science or scientists. Politicians with a healthy scepticism of scientific advocates might simply prefer to arm themselves with this critical set of knowledge.

We are not so naive as to believe that improved policy decisions will automatically follow. We are fully aware that scientific judgement itself is value-laden, and that bias and context are integral to how data are collected and interpreted. What we offer is a simple list of ideas that could help decision-makers to parse how evidence can contribute to a decision, and potentially to avoid undue influence by those with vested interests. The harder part — the social acceptability of different policies — remains in the hands of politicians and the broader political process.

Of course, others will have slightly different lists. Our point is that a wider



BADGER: ANDY ROUSE/NATURE PICTURE LIBRARY; NUCLEAR PLANT: MICHAEL KOHAUT/FILICOR/GETTY; BEE: MICHAEL DURHAM/MINDEN/FLPA

Science and policy have collided on contentious issues such as bee declines, nuclear power and the role of badgers in bovine tuberculosis.

understanding of these 20 concepts by society would be a marked step forward.

Differences and chance cause variation.

The real world varies unpredictably. Science is mostly about discovering what causes the patterns we see. Why is it hotter this decade than last? Why are there more birds in some areas than others? There are many explanations for such trends, so the main challenge of research is teasing apart the importance of the process of interest (for example, the effect of climate change on bird populations) from the innumerable other sources of variation (from widespread changes, such as agricultural intensification and spread of invasive species, to local-scale processes, such as the chance events that determine births and deaths).

No measurement is exact. Practically all measurements have some error. If the measurement process were repeated, one might record a different result. In some cases, the measurement error might be large compared with real differences. Thus, if you are told that the economy grew by 0.13% last month, there is a moderate chance that it may actually have shrunk. Results should be presented with a precision that is appropriate for the associated error, to avoid implying an unjustified degree of accuracy.

Bias is rife. Experimental design or measuring devices may produce atypical results in a given direction. For example, determining voting behaviour by asking people on the street, at home or through the Internet will sample different proportions of the population, and all may give different results. Because studies that report 'statistically significant' results are more likely to be written up and published, the scientific literature tends to give an exaggerated picture of the

magnitude of problems or the effectiveness of solutions. An experiment might be biased by expectations: participants provided with a treatment might assume that they will experience a difference and so might behave differently or report an effect. Researchers collecting the results can be influenced by knowing who received treatment. The ideal experiment is double-blind: neither the participants nor those collecting the data know who received what. This might be straightforward in drug trials, but it is impossible for many social studies. Confirmation bias arises when scientists find evidence for a favoured theory and then become insufficiently critical of their own results, or cease searching for contrary evidence.

Bigger is usually better for sample size.

The average taken from a large number of observations will usually be more informative than the average taken from a smaller number of observations. That is, as we accumulate evidence, our knowledge improves. This is especially important when studies are clouded by substantial amounts of natural variation and measurement error. Thus, the effectiveness of a drug treatment will vary naturally between subjects. Its average efficacy can be more reliably and accurately estimated from a trial with tens of thousands of participants than from one with hundreds.

Correlation does not imply causation. It is tempting to assume that one pattern causes another. However, the correlation might be coincidental, or it might be a result of both patterns being caused by a third factor — a 'confounding' or 'lurking' variable. For example, ecologists at one time believed that poisonous algae were killing fish in estuaries; it turned out that the algae grew where fish died. The algae did not cause the deaths².

Regression to the mean can mislead.

Extreme patterns in data are likely to be, at least in part, anomalies attributable to chance or error. The next count is likely to be less extreme. For example, if speed cameras are placed where there has been a spate of accidents, any reduction in the accident rate cannot be attributed to the camera; a reduction would probably have happened anyway.

Extrapolating beyond the data is risky.

Patterns found within a given range do not necessarily apply outside that range. Thus, it is very difficult to predict the response of ecological systems to climate change, when the rate of change is faster than has been experienced in the evolutionary history of existing species, and when the weather extremes may be entirely new.

Beware the base-rate fallacy. The ability of an imperfect test to identify a condition depends upon the likelihood of that condition occurring (the base rate). For example, a person might have a blood test that is '99% accurate' for a rare disease and test positive, yet they might be unlikely to have the disease. If 10,001 people have the test, of whom just one has the disease, that person will almost certainly have a positive test, but so too will a further 100 people (1%) even though they do not have the disease. This type of calculation is valuable when considering any screening procedure, say for terrorists at airports.

Controls are important. A control group is dealt with in exactly the same way as the experimental group, except that the treatment is not applied. Without a control, it is difficult to determine whether a given treatment really had an effect. The control helps researchers to be reasonably sure that there

are no confounding variables affecting the results. Sometimes people in trials report positive outcomes because of the context or the person providing the treatment, or even the colour of a tablet³. This underlies the importance of comparing outcomes with a control, such as a tablet without the active ingredient (a placebo).

Randomization avoids bias. Experiments should, wherever possible, allocate individuals or groups to interventions randomly. Comparing the educational achievement of children whose parents adopt a health programme with that of children of parents who do not is likely to suffer from bias (for example, better-educated families might be more likely to join the programme). A well-designed experiment would randomly select some parents to receive the programme while others do not.

Seek replication, not pseudoreplication.

Results consistent across many studies, replicated on independent populations, are more likely to be solid. The results of several such experiments may be combined in a systematic review or a meta-analysis to provide an overarching view of the topic with potentially much greater statistical power than any of the individual studies. Applying an intervention to several individuals in a group, say to a class of children, might be misleading because the children will have many features in common other than the intervention. The researchers might make the mistake of 'pseudoreplication' if they generalize from these children to a wider population that does not share the same commonalities. Pseudoreplication leads to unwarranted faith in the results. Pseudoreplication of studies on the abundance of cod in the Grand Banks in Newfoundland, Canada, for example, contributed to the collapse of what was once the largest cod fishery in the world⁴.

Scientists are human. Scientists have a vested interest in promoting their work, often for status and further research funding, although sometimes for direct financial gain. This can lead to selective reporting of results and occasionally, exaggeration. Peer review is not infallible: journal editors might favour positive findings and newsworthiness. Multiple, independent sources of evidence and replication are much more convincing.

Significance is significant. Expressed as *P*, statistical significance is a measure of how likely a result is to occur by chance. Thus *P* = 0.01 means there is a 1-in-100 probability that what looks like an effect of the treatment could have occurred randomly, and in truth there was no effect at all. Typically, scientists report results as significant when the *P*-value of the test is less than 0.05 (1 in 20).

Separate no effect from non-significance.

The lack of a statistically significant result (say a *P*-value > 0.05) does not mean that there was no underlying effect: it means that no effect was detected. A small study may not have the power to detect a real difference. For example, tests of cotton and potato crops that were genetically modified to produce a toxin to protect them from damaging insects suggested that there were no adverse effects on beneficial insects such as pollinators. Yet none of the experiments had large enough sample sizes to detect impacts on beneficial species had there been any⁵.

Effect size matters. Small responses are less likely to be detected. A study with many replicates might result in a statistically significant result but have a small effect size (and so, perhaps, be unimportant). The importance of an effect size

"The question to ask is: 'What am I not being told?'"

is a biological, physical or social question, and not a statistical one. In the 1990s, the editor of the US journal *Epidemiology* asked authors to stop using statistical significance in submitted manuscripts because authors were routinely misinterpreting the meaning of significance tests, resulting in ineffective or misguided recommendations for public-health policy⁶.

Study relevance limits generalizations.

The relevance of a study depends on how much the conditions under which it is done resemble the conditions of the issue under consideration. For example, there are limits to the generalizations that one can make from animal or laboratory experiments to humans.

Feelings influence risk perception. Broadly, risk can be thought of as the likelihood of an event occurring in some time frame, multiplied by the consequences should the event occur. People's risk perception is influenced disproportionately by many things, including the rarity of the event, how much control they believe they have, the adverseness of the outcomes, and whether the risk is voluntarily or not. For example, people in the United States underestimate the risks associated with having a handgun at home by 100-fold, and overestimate the risks of living close to a nuclear reactor by 10-fold⁷.

Dependencies change the risks. It is possible to calculate the consequences of individual events, such as an extreme tide, heavy rainfall and key workers being absent. However, if the events are interrelated, (for example a storm causes a high tide, or heavy rain prevents workers from accessing the site) then the probability of their co-occurrence is much higher than might be expected⁸. The assurance by credit-rating agencies

that groups of subprime mortgages had an exceedingly low risk of defaulting together was a major element in the 2008 collapse of the credit markets.

Data can be dredged or cherry picked.

Evidence can be arranged to support one point of view. To interpret an apparent association between consumption of yoghurt during pregnancy and subsequent asthma in offspring⁹, one would need to know whether the authors set out to test this sole hypothesis, or happened across this finding in a huge data set. By contrast, the evidence for the Higgs boson specifically accounted for how hard researchers had to look for it — the 'look-elsewhere effect'. The question to ask is: 'What am I not being told?'

Extreme measurements may mislead.

Any collation of measures (the effectiveness of a given school, say) will show variability owing to differences in innate ability (teacher competence), plus sampling (children might by chance be an atypical sample with complications), plus bias (the school might be in an area where people are unusually unhealthy), plus measurement error (outcomes might be measured in different ways for different schools). However, the resulting variation is typically interpreted only as differences in innate ability, ignoring the other sources. This becomes problematic with statements describing an extreme outcome ('the pass rate doubled') or comparing the magnitude of the extreme with the mean ('the pass rate in school *x* is three times the national average') or the range ('there is an *x*-fold difference between the highest- and lowest-performing schools'). League tables, in particular, are rarely reliable summaries of performance. ■

William J. Sutherland is professor of conservation biology in the Department of Zoology, University of Cambridge, UK. **David Spiegelhalter** is at the Centre for Mathematical Sciences, University of Cambridge. **Mark Burgman** is at the Centre of Excellence for Biosecurity Risk Analysis, School of Botany, University of Melbourne, Parkville, Australia.
e-mail: wjs32@cam.ac.uk

1. Doubleday, R. & Wilsdon, J. *Nature* **485**, 301–302 (2012).
2. Borsuk, M. E., Stow, C. A. & Reckhow, K. H. *J. Water Res. Plan. Manage.* **129**, 271–282 (2003).
3. Huskisson, E. C. *Br. Med. J.* **4**, 196–200 (1974).
4. Millar, R. B. & Anderson, M. J. *Fish. Res.* **70**, 397–407 (2004).
5. Marvier, M. *Ecol. Appl.* **12**, 1119–1124 (2002).
6. Fidler, F., Cumming, G., Burgman, M., Thomason, N. J. *Socio-Economics* **33**, 615–630 (2004).
7. Fischhoff, B., Slovic, P. & Lichtenstein, S. *Am. Stat.* **36**, 240–255 (1982).
8. Billinton, R. & Allan, R. N. *Reliability Evaluation of Power Systems* (Plenum, 1984).
9. Maslova, E., Halldorsson, T. I., Ström, M., Olsen, S. F. *J. Nutr. Sci.* **1**, e5 (2012).



Writer Aldous Huxley in the late 1930s.

IN RETROSPECT

Brave New World

Philip Ball reconsiders the mix of dystopian science fiction and satire 50 years after Aldous Huxley's death.

When *Brave New World* was published in 1932, science and technology were widely seen as holding utopian promise. The first antibacterials were being developed, the Haber-Bosch process had recently begun

to supply artificial fertilizers, and people were starting to fly between continents and converse across vast distances. Aldous Huxley's bleakly satirical vision of a technocratic, totalitarian state in which the masses are engineered into stupefied contentment

by eugenics, drugs, mindless hedonism and consumerism seemed to scorn that rosy view.

Brave New World
ALDOUS HUXLEY
Chatto & Windus:
1932.

Although it was lauded by some, including the logician and anti-war activist Bertrand Russell, the science boosters felt that Huxley had let the side down. *Nature's* reviewer at the time of publication sniffed that "biology is itself too surprising to be really amusing material for fiction." That reviewer was Charlotte Haldane, whose then husband, the geneticist J. B. S. Haldane, was not averse to predicting the future himself — but in a more optimistic vein.

Gradually, as the star of science waned in the nuclear shadow of Hiroshima and the cold war, *Brave New World* came to be seen as prophetic. But although its status as a classic of twentieth-century literature is rightly secure, what it says about technological development is too often misconstrued.

FEARS FOR THE FUTURE

Huxley's brave new world leaned heavily on the technologies that Haldane had forecast in his essay *Daedalus, or Science and the Future* (1924), particularly the idea of ectogenesis — the gestation of embryos and fetuses in artificial containers. For Haldane, this was a eugenic technique that could improve the human race — as his friend and Aldous's brother, the evolutionary biologist Julian Huxley, also believed. Aldous here, as elsewhere, sided with Russell, who had warned, "I am compelled to fear that science will be used to promote the power of dominant groups, rather than to make men happy." In a 1932 article, biochemist and Sinophile Joseph Needham described *Brave New World* as a note-perfect realization of Russell's concerns.

But Huxley's dystopia upset some champions of scientific progress much more than it did Charlotte Haldane. H. G. Wells, whose 1923 novel *Men Like Gods* served up a characteristically glorious scientific utopia, felt personally offended, allegedly saying "a writer of the standing of Aldous Huxley has no right to betray the future as he did in that book" (Huxley admitted that irritation with Wells's book was partly what provoked him to write *Brave New World* in the first place.)

So *Brave New World* did not appear out of nowhere, but was a contribution to a vigorous interwar debate about the influence of science on society, not least the roles of reproductive technologies. That debate was exemplified by the *To-day and To-morrow* essay series — of which *Daedalus* was the first — published in Britain by Kegan Paul between 1923 and 1931. Through it, scientists, philosophers, politicians, artists and feminists engaged deeply in a conversation that has never since been matched.

Within that context, *Brave New World* can be read as a turning of the tide in terms of perceptions of what science would bring: from optimism to foreboding. With the benefit of perspective, what should we make of it now?

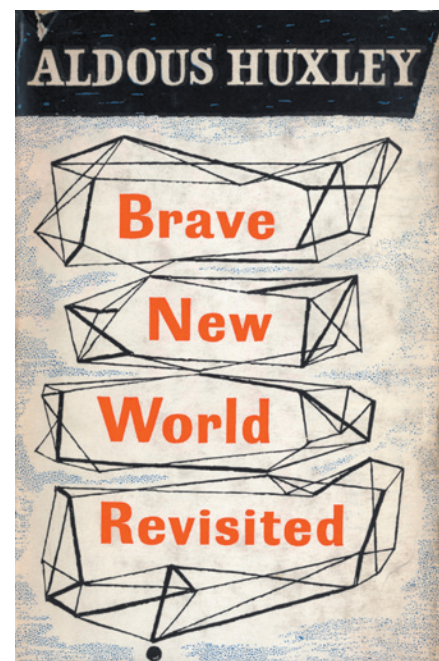
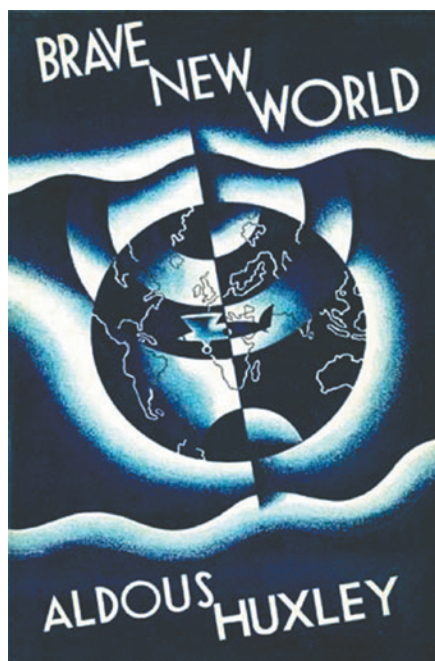
The story is set in AD 2540 (or 632 'After Ford', the god of mass production). A World State manufactures its citizens by growing fetuses in bottles according to "Bokanovsky's Process": cloning many embryos from a single fertilized egg and treating them with chemical agents during development to produce a five-tier caste system of intelligence. Sex is recreational, love is obsolete and the idea of family is obscene.

Outside this society live small communities of 'savages' who maintain the old ways of reproduction and religion. One of them, a young man called John, has become eloquent (rather too much so, Huxley admitted) by reading Shakespeare — hence the quote from *The Tempest* that gives the book its ironic title. John echoes Miranda's naive phrase as he initially thrills to the prospect of visiting civilization, and then is horrified by the shallow, hedonistic passivity of its citizens. Lacking art, religion and any sort of genuine passion or curiosity, this stagnant society has, John says, paid "a fairly high price" for its empty happiness. He is eventually driven to despair and suicide.

The book begins with its most famous set-piece: the human 'hatchery'. Decked out in the "glass and bleakly shining porcelain of a laboratory", it houses incubators that contain "racks upon racks of numbered test-tubes". Thus, *Brave New World* reimagines the old myth of making artificial people (anthropoeia) in a form that was appropriate for the early twentieth century: no longer a lone and secretive quasi-chemical pursuit, but an industrial-scale operation. This is a perceptive revision of Mary Shelley's *Frankenstein* (1818), although it was anticipated in Karel Čapek's 1921 play *R.U.R.*, which described the manufacture of flesh-and-blood 'robots'.

In literary terms, Huxley's satire is rich, but his story and characters are thin. This is a common feature of science fiction from Jules Verne to J. G. Ballard, and has led some critics to insist that the genre can never produce 'true literature'. That is to utterly miss its point. As Robert Philmus argued in *Into the Unknown* (1970), science fiction from Jonathan Swift's 1726 work *Gulliver's Travels* onwards "draws upon the metaphors inherent in current ideas and transforms them into myth". Myth demands sketchy characters — it has concerns beyond the

"The book's lasting power is as a tale about ways in which we can lose our humanity."



Aldous Huxley's 1958 essay *Brave New World Revisited* assessed the dystopian vision of his 1932 classic.

modernist focus on the individual psyche. Often those concerns are satirical: by materializing ideas, their limitations are revealed. As with Swift, so with Huxley.

In other words, *Brave New World*, like most classics of science fiction, is less a work of invention than one of analysis — it is about the present (in this case, the period between the wars), not the future. Huxley's target was contemporary fears of totalitarian communism and fascism, wariness about eugenics and scientific triumphalism, and anxieties about consumerism ("Our Ford" is the profanity of choice) and mass docility. He hits all these targets with humour that has true bite. The real issue is broader than the details — as Huxley put it, "not the advancement of science as such [but] the advancement of science as it affects human individuals".

MEANING MISREAD

What irks me is how persistently the book is misread as foresight, often for rhetorical and dogmatic purposes. When Louise Brown, the first baby to be born through *in vitro* fertilization (IVF), arrived in 1978, *Newsweek* trumpeted her first "lusty yell" as "a cry heard round the brave new world". The spectre of mass-produced, 'dehumanized' citizens was brandished by bioethicist Leon Kass, from his early opposition to IVF through to his thwarting of stem-cell research as the head of George W. Bush's Council on Bioethics.

NATURE.COM
For a podcast of Philip Ball on *Brave New World*, see: go.nature.com/hoozdl

Brave New World has been co-opted as an off-the-shelf apocalyptic warning about where all such advances, from

xenotransplantation (the interspecies grafting or transplanting of organs and tissues) to cloning, will lead.

All the same, one has to admit that Huxley's vision was sometimes right on the money. His state controls its citizens not by Orwellian repression but through a drug (soma) administered to engender bovine passivity, along with the opiate of consumerism. "A really efficient totalitarian state would be one in which [leaders] control a population of slaves who do not have to be coerced, because they love their servitude," Huxley wrote. In his 1958 essay *Brave New World Revisited*, he rightly noted that "it now looks as though the odds were more in favour of something like *Brave New World* than of something like 1984". His dystopian state uses non-stop, trivial, sensual distractions to prevent people from paying too much attention to social and political realities. One doesn't have to be a conspiracy theorist to see those enervating distractions — infotainment, social media, celebrity-dominated news — being useful today to both authoritarian and liberal regimes.

Yet despite such flashes of prescience, *Brave New World* is not a cautionary fable about particular trajectories in science or politics. The Central Hatchery is not prophetic; it is symbolic. Like *Frankenstein*, the book's lasting power is as a tale about ways in which we can lose our humanity. These ways differ in every age, but the result is much the same. ■

Philip Ball is a writer based in London and the author of *Unnatural: The Heretical Idea of Making People*.
e-mail: p.ball@btinternet.com



DOGWOOF

Adi Zulkadry (left) and Anwar Congo, who between them killed hundreds of fellow Indonesians, are made up as their victims in *The Act of Killing*.

ANTHROPOLOGY

The science of impunity

Tanguy Chouard rates a film probing the psyches of Indonesia's paramilitary killers.

In 1965–66, between half a million and 3 million 'communists' — union members, landless farmers, intellectuals and ethnic Chinese people — were exterminated by paramilitary gangsters sponsored by the new military dictatorship of Indonesia. The same regime has been in power and persecuting its opponents ever since.

Twelve years ago, film director Joshua Oppenheimer realized that the killings had received direct support from the West but were hardly ever researched. To him, the situation constituted a living experiment in mass impunity, as if "the Nazis were still in power". Having lost relatives to the Holocaust, he felt he owed it to the victims and their families to document the genocide.

He could not film the survivors without compromising their safety, so Oppenheimer turned to the perpetrators, who boasted about their 'heroic past' (see go.nature.com/2b6v7v). After interviewing more than 40 death-squad leaders, Oppenheimer earned a PhD in documentary film-making and secured academic funding to explore more deeply how the killers and their government sponsors viewed themselves, and how they wanted to be seen — and feared — by Indonesian society.

He formulated an outlandish anthropology project, turning the camera over to the killers and inviting them to dramatize their

story as they wished. The result is *The Act of Killing*, a fiercely original experiment in documentary film-making that exposes the entrails of a brutal regime of impunity. The film's diffusion throughout Indonesia is shaking the country's bedrock of violence.

The Act of Killing is "not a movie about the past", Oppenheimer insists; "it is a movie of the imagination". And the perpetrators' imaginations, fuelled by Hollywood gangster movies, emerge as a surreal hotchpotch of crime scenes, involving a cowboy, a drag queen, macaques, elephants and a stuffed crocodile. The grotesque gives way to the disturbing, and bright colours fade to darkness, in a chilling crescendo of realism. At the film's centre are the persistent nightmares of Anwar Congo, a man believed to have executed about 1,000 people.

After impersonating a victim, the smartly dressed, often avuncular Congo briefly empathizes with those whom he tortured. But in one of the most terrifying scenes, he suddenly pulls out his knife and lacerates a teddy bear, a stand-in for the baby of a 'communist' under interrogation, declaring: "this is what we do to those who bribe us with their children."

The Act of Killing
DIRECTED BY JOSHUA
OPPENHEIMER
UK DVD release:
25 November 2013.

Fellow executioner Adi Zulkadry is remorseless. "It's all about finding the right excuse," he lectures, as fake blood is plastered onto his face. When Oppenheimer confronts him about the Geneva Conventions, Zulkadry retorts: "War crimes are defined by the winners. I'm a winner, so I can make my own definition." Zulkadry is more lucid than his fellow film-makers about the consequences of exposing truth. "For me, reopening this case is a provocation to fight," he warns.

Aware that a public banning of the film would justify official violence against anyone seeing it, Oppenheimer first gave private screenings to Indonesian journalists, celebrities and human-rights activists. Huge media coverage ensued and the film, now freely downloadable, has been seen by more than 200,000 Indonesians, he says.

Oppenheimer's freakish project allowed him to navigate the murky waters of Indonesia's 'democracy' and then to jolt its collective conscience into ending 47 years of media silence. Impunity, in retrospect, was a metastable state — a regime that most knew was wrong, but that no one could fully expose without a powerful catalyst. ■

Tanguy Chouard is a senior biology editor at Nature.

Q&A Peter Westwick

Surfing scientist

Historian Peter Westwick and his colleague Peter Neushul thought up their scientific history of surfing, *The World in the Curl* (Crown, 2013), on boards off the coast of California. As the winter surfing season gets into full swing, Westwick talks about warfare, wetsuits, climate change and forecasting surf.

ILLUSTRATION BY NICK HIGGINS

What does surfing reveal about our relationship with nature?

Surfing is often seen as a romantic retreat to the wild ocean among seals and dolphins — finding yourself no longer at the apex of the food chain. But in *The World in the Curl*, I and my co-author and fellow historian, Peter Neushul, are trying to show that surfing is caught up with industry, technology and commerce. In the morning I check conditions on my laptop, then paddle out in a neoprene wetsuit on an ultralight board. The technology connects us to nature but also changes our relationship with it.

How has science influenced the development of the sport?

The popularization of surfing over the past century is linked to the evolution of surfboard design. Early surfers in Hawaii used giant redwood planks. To drag a 50-kilogram chunk of wood across the beach, then wrestle it through walls of white water, you had to be a phenomenal athlete. These days you can get a 2.5-kg board made of polyurethane foam, fibreglass and resin. There was early experimentation with balsa wood, which is light until it absorbs water and sinks. In 1928, surfer Tom Blake devised a hollow wooden surfboard that was probably inspired by the wing of the Lockheed Vega aeroplane. But the real revolution came from synthetic materials made during wartime.

Who designed the modern surfboard?

Studying mechanical engineering at the California Institute of Technology [Caltech] in the early 1940s, Robert Simmons ran across polystyrene foam and polyester resin, then mass-produced for aviation. With his knowledge of water flow, connected to Caltech's work on air-dropped torpedoes, he designed streamlined boards. His 'hydrodynamic planing hull' soon became the standard. In the 1970s, aerospace engineer Tom Morey, who had worked on rocket nozzles, invented the boogie board, a simple foam panel that got millions of people riding waves. A recent backlash against new materials has seen some surfers return to solid wooden boards that promise an unmediated encounter with the wave.



How did the wetsuit come about?

During the Second World War, Allied divers who defused underwater mines wore drysuits for warmth, but air trapped inside the suits caused them to wrinkle and pinch. American physicist Hugh Bradner, who worked on the atomic bomb as part of the Manhattan Project, had a counter-intuitive insight: you don't have to stay dry to stay

warm. His suit of neoprene, a synthetic material that was developed by DuPont in the 1930s as a rubber substitute, insulated divers with a layer of trapped water. Neoprene is also a shock absorber and protected divers from underwater explosions. With a wetsuit you can surf in the winter off California — even off Alaska and Antarctica.

When did the modern age of surf forecasting begin?

Again, it began with the Second World War. Allied strategy involved moving armies from ship to shore. Landing craft capsizing in the surf zone could change the course of battles. Military planners realized that you can't launch an amphibious invasion when the waves are big, so the size of waves became an issue. In 1941, oceanographer Walter Munk began to work on the scientific problem of how to define and measure ocean waves. He found that by measuring the speed and direction of winds in the middle of ocean, you could predict how big waves would be on beaches thousands of miles away a few days later. His theory helped the Allies to land at Normandy on D-Day.

What tools are used to predict waves today?

The basic premise of surf forecasting has not changed much. But in recent decades there have been great advances in how we collect the data. Electronic buoys along the whole Pacific coast of the United States measure wind and swell height. Satellites can tell the speed, size and direction of storms in the vast expanse of the Pacific. Then supercomputers crunch the data and calculate the height, frequency and direction of waves resulting from the storm, so you can predict the surf on a given beach several days later.

Can waves be engineered?

On many beaches this happens already, if usually unintentionally, because harbours, piers and sea walls change wave patterns. There have been attempts to build artificial reefs for surfing in California, the United Kingdom, New Zealand and India. These have mostly failed. Engineers have tried to create surf outside the ocean, for instance with the FlowRider, which makes a small stationary wave by propelling water against a curved sheet of foam.

Will global warming change surfing?

Surfers are at the frontline of environmental change. More severe storms will make the surf more extreme and less consistent. Rising seas will change where people can surf. I'm not sure that the surfing community has entirely woken up to these facts yet. It's been a long struggle to convince surfers to work to preserve their own environment. ■

INTERVIEW BY JASCHA HOFFMAN



SHANNON STENT/GETTY

Correspondence

MOOCs taken by educated few

Massive open online courses (MOOCs) have been hailed as an educational revolution that has the potential to override borders, race, gender, class and income (see go.nature.com/hanoau). However, a survey of active MOOC users in more than 200 countries and territories has revealed that most students on these courses are already well educated — and that they are predominantly young males seeking to advance their careers.

Our data are drawn from 34,779 responses to a July 2013 survey by the University of Pennsylvania, USA, of participants in 32 course sessions of the online education service Coursera (see <https://www.coursera.org/penn>). We found that 83% of surveyed students already had a two- or four-year post-secondary degree (see 'MOOCs are not reaching the disadvantaged', red bars), with 44.2% reporting education beyond a bachelor's degree (see go.nature.com/cvjp8u).

Furthermore, the prior educational standard among MOOC students across the world far exceeds that of the general population in their own countries (see figure, blue bars; source: www.barrolee.com).

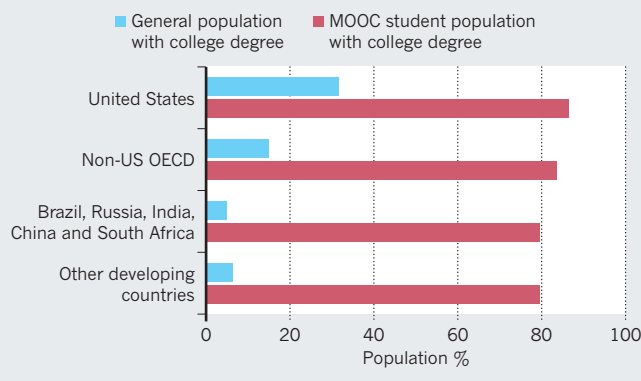
This educational disparity is particularly stark in Brazil, Russia, India, China and South Africa, all of which are prime candidates for MOOC education. In those countries, almost 80% of MOOC students come from the wealthiest and most well-educated 6% of the population.

We found that men account for 56.9% of all MOOC students (and 64% in countries outside the Organisation for Economic Co-operation and Development; OECD). Also, almost 70% of MOOC students are already in employment (these data are not shown).

Far from realizing the high

MOOCs ARE NOT REACHING THE DISADVANTAGED

The majority of students on massive open online courses (MOOCs) are already well educated compared with the general population.



ideals of their advocates, MOOCs seem to be reinforcing the advantages of the 'haves' rather than educating the 'have-nots'. Better access to technology and improved basic education are needed worldwide before MOOCs can genuinely live up to their promise.

Ezekiel J. Emanuel *University of Pennsylvania, Philadelphia, USA.*
vp-global@upenn.edu

**On behalf of 6 co-signatories (see go.nature.com/8lqpa5 for a full list).*

Backing up forensic DNA evidence

Pakistan's leading Islamic guidance body, the Council of Islamic Ideology, recently declared that DNA profiling is inadequate as primary evidence for rape crimes and should be supported by other forms of evidence laid out in Islamic law — for example, a confession, or confirmation from four adult male eyewitnesses (see go.nature.com/myl1da).

The council's ruling is based on the inability of DNA testing to distinguish between forced and consensual sex. We hope, however, that it will be sufficiently flexible to accommodate DNA testing when it could be decisive — for example, in cases of child abuse, when geographical location

precludes the presence of eyewitnesses, or when the rape victim is murdered.

Collaboration between forensic scientists and religious scholars would foster such flexibility by improving mutual understanding and facilitating more informed decision-making.

Rape crimes in Pakistan are currently not accountable to civil law, but are judged according to the 1979 Hudood Ordinance, which faces mounting criticism from human-rights organizations and moderate sectors of civil society. By diminishing the importance of DNA testing, the council seems to be endorsing the stance of the Hudood Ordinance.

Mushtaq Hussain, Ammara Mushtaq *Dow University of Health Sciences, Karachi, Pakistan.*
mushtaq.hussain@duhs.edu.pk

Smarten up on intelligence genetics

You write that "50% of variability in intelligence seems to be inherited" (*Nature* 502, 26–28; 2013). This figure is derived from quantitative genetic studies that do not seem to be founded on sound scientific reasoning.

It is unlikely that quantitative genetics can be reasonably applied to mental traits in

humans. There can be no single value, or even range of values, for the heritability of intelligence, because environmental differences vary vastly between populations — consequently, published values for the heritability of intelligence range between 0% and 100%.

Indeed, the heritability of intelligence has little to do with its malleability — so why estimate it in the first place? The whole idea seems brain-dead to me.

M. Velden *Department of Psychology, University of Mainz, Germany.*

Database differences not citation errors

Differences in citation records across international databases reflect variations in their coverage of the scientific literature, rather than inaccuracies (D. Shotton *Nature* 502, 295–297; 2013).

Google Scholar, for example, indicates the number of citations that have appeared in almost any online scientific material. Thomson Reuters' Web of Science records only the number of citations in selected high-impact journals. Both numbers are accurate on the basis of each source's indexing criteria, so each should be interpreted individually.

Mohammad H. Nowroozadeh *Poostchi Eye Research Centre, Shiraz University of Medical Sciences, Shiraz, Iran.*
noroozadeh@gmail.com

CONTRIBUTIONS

Correspondence may be sent to correspondence@nature.com after consulting the author guidelines at <http://go.nature.com/cmchno>. Alternatively, readers may comment online: www.nature.com/nature.

EXTRASOLAR PLANETS

An infernal Earth

Orbiting less than two stellar radii above the visible surface of a Sun-like star, the Earth-sized exoplanet Kepler-78b is a hellish world. But its existence bodes well for the discovery and characterization of habitable planets. [SEE LETTERS P.377 & P.381](#)

DRAKE DEMING

A prime goal of exoplanetary science — the study of planets beyond the Solar System — is to find and characterize Earth-like planets orbiting Sun-like stars. This is a daunting task because, in relation to the cosmos, Earth is meagre in both size and mass, so Earth-like planets would be easily lost in the glare of a Sun-like star. A giant step towards finding another Earth was taken in 1995, when the first extrasolar planets were discovered orbiting Sun-like stars¹. But those exoplanets were hydrogen-dominated gas giants in scorching orbital zones, as unlike Earth as could be imagined. Fortunately, NASA's Kepler space mission has changed scientists' perspective on exoplanets. And now, in two papers in this issue, Howard *et al.*² (page 381) and Pepe *et al.*³ (page 377) independently report measurements of an exoplanet, Kepler-78b, showing conclusively that the planet's mass is about 80% greater than Earth's, and that its radius is only 20% greater — a virtual twin of Earth by astronomical standards.

Kepler has found thousands of rocky and icy exoplanets whose orbits cause them to periodically block the light of their parent stars. The amount of stellar light that they block provides an estimate of their size, and Kepler has revealed that planets comparable in size to Earth are abundant in our Galaxy⁴. Although Kepler measures the size of exoplanets with exquisite precision, the planets' composition has been much more difficult to determine. Knowing the bulk composition of an exoplanet requires a determination of its mass using ultra-high-precision Doppler spectroscopy, which measures shifts in the wavelength of the parent star's light caused by its reflex motion as the planet orbits it. Unfortunately, most of the planets discovered by Kepler produce Doppler shifts that are too small to be measured. But for Kepler-78b, Howard *et al.* and Pepe *et al.* describe accurate mass measurements that allowed them to derive an average density for the planet that is almost identical to Earth's 5.5 grams per cubic centimetre.

The measurement of Kepler-78b's mass was made possible by its close proximity to its host star, which greatly increases the star's Doppler reflex. But that boost of the Doppler signal

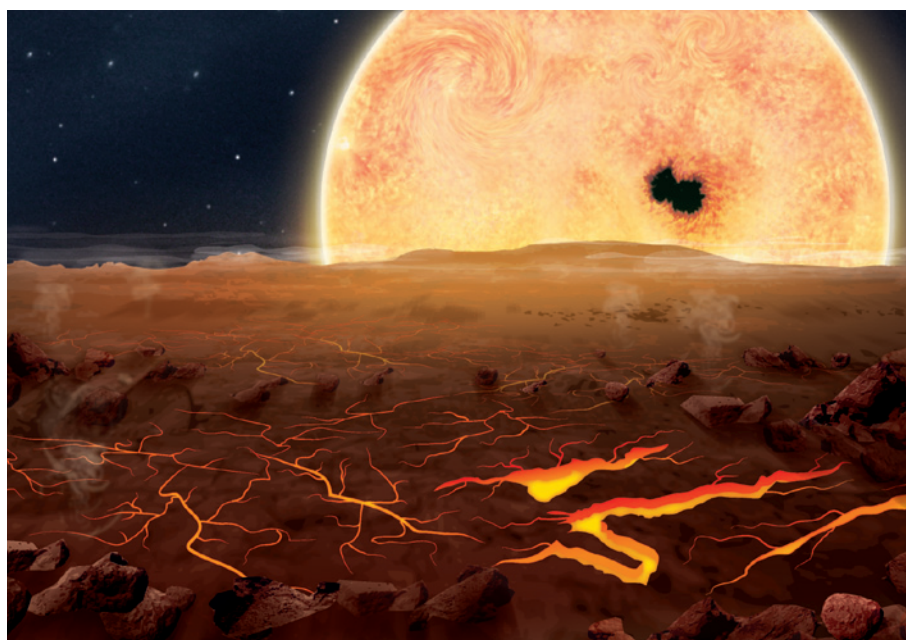


Figure 1 | Exoplanet Kepler-78b. This artist's impression of Kepler-78b shows the view from the planet's surface, with the disk of its host star filling much of the sky.

comes at the price of a hellish environment. The planet's orbit lies less than two stellar radii above the star's visible surface, and a view from the surface of Kepler-78b would be dominated by the blazing disk of the star, filling about half of the sky from horizon to zenith (Fig. 1). According to current understanding, the chances of life in such an environment are nil. Nevertheless, Kepler-78b is an encouraging sign in the search for extrasolar habitable worlds.

The density of the planet indicates that it is probably composed of rock and iron, very much like Earth. How it came to reside in its current 8.5-hour orbit is uncertain. Among the more exotic possibilities is that it is the remnant core of a disrupted gas giant⁵. Regardless of its history since it formed, Kepler-78b probably originated by a process of accretion in a protoplanetary disk of gas and dust, and it shares that origin with Earth. But many other aspects of Earth seem to be unique, raising the question of whether we can reasonably expect to find similar worlds that host exoplanetary life forms. The existence of Kepler-78b shows that, at the very least,

extrasolar planets of Earth-like composition are not rare.

If, as seems reasonable, planets of Earth-like composition are not uncommon in the Milky Way, it should be possible to find one that is both nearby in cosmic terms and exhibits the favourable orbital geometry that blocks the light from its star as it orbits — a key feature enabling characterization. NASA is currently preparing the Transiting Exoplanet Survey Satellite (TESS) to search the entire sky for such favourable exoplanets. For the best cases found, we will turn to mass measurements by Doppler spectroscopy, and characterization of the exoplanetary atmosphere using the James Webb Space Telescope (JWST), which is planned for launch in 2018. As with Kepler-78b, the scientific yield of the JWST and TESS will be greatly amplified by the use of ultra-high-precision Doppler spectroscopy to measure exoplanetary masses. This technique has continued to advance in sensitivity, so that measurements once only dreamed about — namely, precision at the level of 1 metre per second in reflex velocity — are becoming routine.

The main instruments enabling Doppler

JASIEK KRZYSZTOFIK/NATURE

spectroscopy to be used to weigh exoplanets have been the High Resolution Echelle Spectrometer (HIRES) on the Keck telescope⁶, with which Howard *et al.* made their observations, and the High Accuracy Radial velocity Planet Searcher (HARPS) at the European Southern Observatory's 3.6-metre telescope in La Silla, Chile⁷. HARPS has been particularly successful for these demanding measurements because it was designed for exactly this purpose. A Northern Hemisphere version, HARPS-N, became operational in 2012 on the 3.57-metre Telescopio Nazionale Galileo at the Roque de los Muchachos Observatory in La Palma, Spain, and has made a spectacular debut by enabling Pepe *et al.* to measure the mass of Kepler-78b.

If applied to exo-Earths that TESS discovers, HARPS-N and HIRES will produce mass measurements for exoplanets whose environments are more temperate than that of Kepler-78b. By focusing particularly on small stars cooler than the Sun, TESS should find exo-Earths whose mass can be measured by trading the close-in orbit of Kepler-78b for more distant orbits around low-mass stars, approaching orbital zones where life is possible. That trade-off probably cannot be pushed to the point of measuring an Earth twin orbiting once per year around a Sun twin, but it will allow future scientific teams to probe habitable planets orbiting small stars. Kepler-78b thereby foreshadows leaps forward

in the search for life beyond the Solar System. ■

Drake Deming is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.
e-mail: ddeming@astro.umd.edu

1. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
2. Howard, A. W. *et al.* *Nature* **503**, 381–384 (2013).
3. Pepe, F. *et al.* *Nature* **503**, 377–380 (2013).
4. Fressin, F. *et al.* *Astrophys. J.* **766**, 81 (2013).
5. Jackson, B., Stark, C. C., Adams, E. R., Chambers, J. & Deming, D. Preprint at <http://arxiv.org/abs/1308.1379> (2013).
6. Vogt, S. S. *et al.* *Proc. SPIE* **2198**, 362 (1994).
7. Mayor, M. *et al.* *ESO Messenger* **114**, 20–24 (2003).

This article was published online on 30 October 2013.

PLANT BIOMECHANICS

High-endurance algae

Breaking waves place repeated loading on marine algae, which can lead to death by fatigue. But observations of one alga suggest that its joint structure, which lacks transverse connections, confers fatigue resistance.

EMILY CARRINGTON

Rocky shores are pounded by surf, with each wave delivering a fresh assault on the attached plants and animals about once every ten seconds — more than 8,000 times a day or almost 3 million times a year. Most organisms are simply not up to the task of surviving in this environment; only a select few are able to endure these high-intensity workouts and proliferate. Writing in the *Journal of Experimental Biology*, Denny *et al.*¹ reveal a key design feature of one of the most successful surf-zone competitors: the strong, fatigue-resistant joints of coralline algae.

The authors focused on an alga that is common to the waviest coasts of California: *Calliarthron cheilosporioides*, a beautiful branched plant the size of your hand. Each branch of this red alga resembles a necklace of pink beads, with a pale decalcified joint (geniculum) connecting each calcified bead (intergeniculum) to the next. The authors knew that these numerous joints conferred flexibility on what would otherwise be a rigid structure, allowing the plant to sway to and fro and reduce the impact of large breaking waves. But they also knew, from previous studies², that most flexible algae are constructed of tissues that are prone to fatigue arising from accumulated progressive and localized structural damage caused by repeated loading from wave forces. This damage comes in the form of microcracks that concentrate stress and then elongate and propagate catastrophically through the material — a process elegantly described in 1921 by the engineer Alan Arnold Griffith³.

As a result, many algae get weaker with each passing wave and die from fatigue sooner than would be expected on the basis of their nominal strength.

Calliarthron, however, has a lifespan of up to six years, which is relatively long for a surf-zone plant. Could it be resistant to fatigue? One

clue lies in the microstructure of the genicular joint. Denny and colleagues observed that the elongated cells of the joint are arranged in a single tier that runs parallel to the axis of growth, with each joint cell terminating at one end of a calcified bead (Fig. 1a). They proposed that because these joints are not connected transversely to each other, there is no structure to concentrate stress and propagate cracks from one cell to the next.

A simplified illustration of this concept is given by a dancing marionette. The puppet's body and appendages are suspended by strings, each with a direct attachment to a control bar above. As the puppeteer manipulates the bar, force is transmitted through the strings to make the puppet dance. If one string snaps, the arm or leg it supported will fall lifeless,

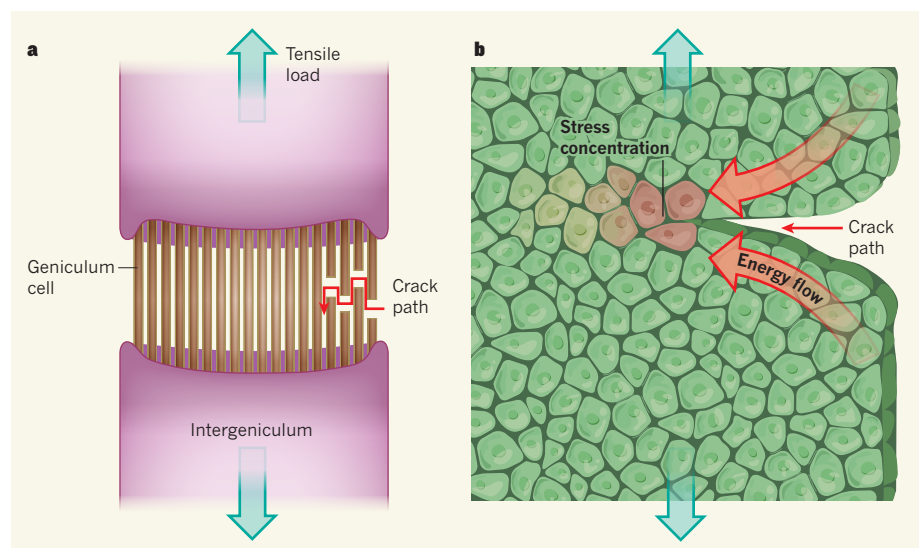


Figure 1 | No stress. **a**, The tissue of the alga *Calliarthron cheilosporioides* comprises flexible genicular joints connecting calcified intergeniculum regions. Each geniculum contains a single tier of elongated cells that lie parallel to the axis of growth. Denny *et al.*¹ show that this cellular structure helps the alga to withstand the repeated tensile loading of waves — the lack of transverse connections between the geniculum cells means that when one cell breaks, strain energy is not passed to the next cell. Furthermore, stress does not concentrate at the crack tip, so the crack path will 'meander' through the tissue and not propagate readily. **b**, The tissues of other algae more closely resemble a homogeneous material of interconnected isodiametric cells. Once a crack path forms in this material, strain energy can flow more easily towards the tip of the crack, allowing stress to concentrate so that the crack propagates in a more directed way, perpendicular to the axis of loading.

but the rest of the doll will continue to follow the puppeteer's guidance — the failure of one string has little impact on the performance of the others. The key is that the strings are not connected directly to each other and therefore act independently.

Denny *et al.* hypothesized that the genicular joint material of *Calliarthron* is not homogeneous, but instead comprises a bundle of parallel cells acting as independent cables. They sought indirect support for this view by comparing the stiffness of a joint in tension with its stiffness in shear (denoted E and G , respectively, and measured in pascals). Materials-science theory establishes a ratio E/G of 3 for a homogeneous material; the loss of transverse connections between cells reduces shear stiffness and causes E/G to be much greater than 3. Denny and colleagues' experimental data showed that the E/G ratio for *Calliarthron* is more than 10, confirming that each joint acts as a bundle of strong, extensible, loosely connected parallel cables. This suggested that the alga's structure resists crack propagation and fatigue.

The authors next directly measured fatigue

in *Calliarthron* by placing the alga in a custom-made device that mimicked repeated loading by waves. When loaded to 60% of its nominal strength, the alga survived more than ten million cycles, the equivalent of more than three years of waves crashing on it every ten seconds. Because the majority of waves encountered by *Calliarthron* produce lower forces, the predicted longevity of this alga is much longer than its observed lifespan of six years. In short, the authors conclude that failure by fatigue is not likely in *Calliarthron*, and that only rare, extreme waves that exceed the nominal strength of the joints would cause death.

Most of the macroalgae that compete with *Calliarthron* for space and light have a tissue construction that is more prone to crack propagation (Fig. 1b). Over time, repeated loading by the surf gradually prunes back new tissue growth. The net result is a population of algae that are much smaller — and therefore less competitive — than would grow according to estimates based on the nominal strength of the algal tissue². By contrast, the crack-stopping tissue construction of *Calliarthron* joints

confers a competitive edge by minimizing pruning by waves.

The single tier of genicular cells is the feature of the joints of *Calliarthron* that confers fatigue resistance. The authors point out that the joints of other lineages of coralline algae are constructed differently, with multiple tiers that may not act independently in shear, and therefore might be less resistant to fatigue. Although this evolutionary story awaits further investigation, we now know how *Calliarthron*, with a joint structure that resists crack propagation, is able to sway back and forth indefatigably to the beat of the surf. ■

Emily Carrington is at the Friday Harbor Laboratories, University of Washington, Friday Harbor, Washington 98250, USA.
e-mail: ecarring@u.washington.edu

1. Denny, M., Mach, K., Tepler, S. & Martone, P. J. *Exp. Biol.* **216**, 3772–3780 (2013).
2. Mach, K. J., Tepler, S. K., Staaf, A. V., Bohnhoff, J. C. & Denny, M. W. *J. Exp. Biol.* **214**, 1571–1585 (2011).
3. Griffith, A. A. *Phil. Trans. R. Soc. Lond. A* **221**, 163–198 (1921).

ivers and lakes are supersaturated with CO₂ — that is, the measured concentration in river water usually exceeds the equilibrium value for the exchange between CO₂ present as a gas in the atmosphere and that present as a dissolved substance in water⁴. This excess concentration in water drives molecular diffusion across the water–air interface. The available measurements of CO₂ in surface waters cover North America, parts of Europe, South Africa and Japan, but the data are still sparse for Africa and Asia, which adds significant uncertainty to the authors' global analysis.

To obtain regional gas-transfer velocities, the researchers analysed the results of published tracer experiments⁵ of gas-exchange rates across the air–water interface, which are related to the level of turbulence. For lakes, this gas-exchange parameter correlates with lake size and the average wind field. In rivers, the gas-transfer velocity is rapid in steep terrain — in the headwaters — and slower towards the lowlands.

To estimate the global surface area of river systems, Raymond *et al.* made use of high-resolution geographical data obtained from space-shuttle missions, detailed river monitoring in the United States, and climate parameters from the COSCAT database (which collects information about river catchment areas globally) to derive statistical correlations between surface area and climatic parameters such as precipitation and temperature. They also revised an earlier census of lakes and reservoirs⁶.

This global analysis reveals an annual CO₂ flux of 1.8 petagrams (Pg; 1 petagram is 10⁹ tonnes) from rivers to the atmosphere, and 0.32 Pg from lakes and reservoirs. The study

BIOGEOCHEMISTRY

Conduits of the carbon cycle

Emissions of carbon dioxide from inland waters to the atmosphere are a crucial link in the global carbon cycle. A comprehensive analysis reveals that this connection is much stronger than was previously thought. [SEE ARTICLE P.355](#)

BERNHARD WEHRLI

The global river system acts as a gigantic transportation network for water and dissolved substances, but this pipeline is leaking carbon dioxide to the atmosphere at surprisingly high rates. On page 355 of this issue, Raymond *et al.*¹ report that rivers emit about five times more CO₂ to the atmosphere than all lakes and reservoirs put together. The authors' spatial analysis reveals a flux of this greenhouse gas that is larger than previously estimated and dominated by hotspot regions in the tropics.

Only about half of anthropogenic CO₂ emissions accumulate in the atmosphere²; their effects on climate are being mitigated by mechanisms that bring about uptake by the oceans and by terrestrial vegetation. To promote carbon sequestration on land, we need to know how and where these large amounts of carbon are removed from the atmosphere.

Climate scientists estimate the strength of carbon sinks on land by running global circulation models of the atmosphere, and by

identifying regional sinks using monitoring stations that measure the global distribution and variability of atmospheric CO₂ concentrations. This top-down approach is limited by the spatial resolution of both the models and the data. By contrast, ecosystem scientists approach the problem from the bottom up: they measure the CO₂ uptake and release rates of different natural and agricultural vegetation systems. The large spatial and seasonal variability of photosynthesis and respiration poses a significant challenge in scaling up these local observations. The lateral export of carbon from land to river networks is a complicating factor in determining regional carbon budgets^{2,3}, but solving this problem also provides the opportunity to monitor and model this important flux with high spatial resolution.

To calculate global estimates of CO₂ emissions from inland waters, Raymond *et al.* created spatially resolved data sets of three parameters: CO₂ concentrations in surface waters; the velocity of gas transfer to the atmosphere; and the surface areas of rivers and lakes. It has long been known that most

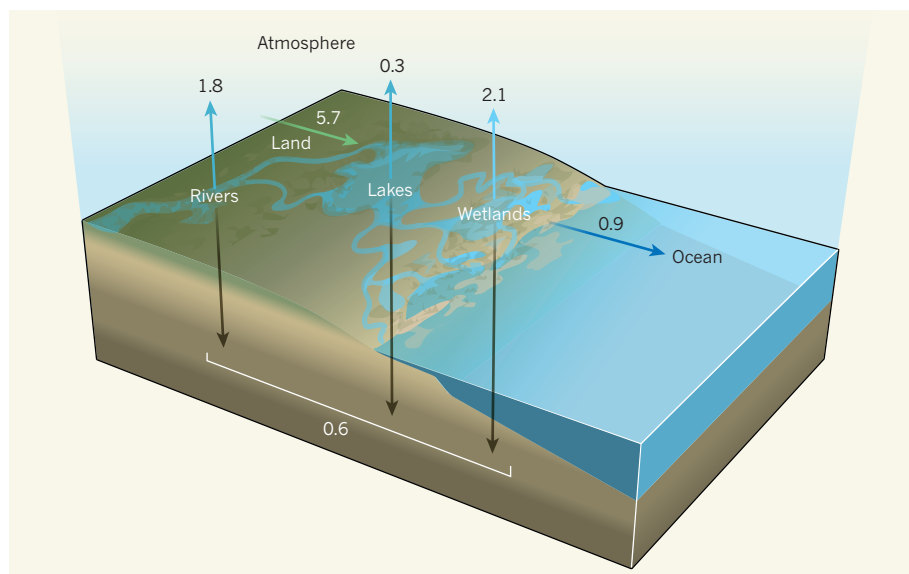


Figure 1 | Inland waters in the global carbon cycle. The numbers above the rising arrows indicate carbon dioxide emissions from inland waters in petagrams of carbon per year (1 Pg is 10^9 tonnes), including data for rivers and lakes now reported by Raymond *et al.*¹ and for emissions from wetlands³. Also shown are data for the export of dissolved carbon by rivers to the ocean² (dark blue arrow) and the amount of organic carbon stored in sediments and wetland soils² (descending arrows). The numbers indicate that a substantial fraction of the carbon fixed by terrestrial vegetation must be laterally exported from land into surface waters (green arrow), which affects regional carbon budgets on land. (Figure adapted from ref. 3.)

excluded wetlands, for which coarser estimates are available³. The resulting carbon transfer from terrestrial systems to inland waters therefore amounts to 5.7 Pg per year^{2,3}, substantially higher than was previously thought (Fig. 1).

Where is all this carbon coming from? There are three main possibilities: soil respiration, which increases inorganic carbon concentrations in groundwater; soil erosion, which transports organic-rich particles into streams; and the entry of a substantial amount of dead biomass into water courses in forested and wetland systems⁷. At present, it is not clear which source contributes most to carbon transfer from the land to surface waters. Researchers are therefore using advanced methods of organic and isotope geochemistry to disentangle these terrestrial carbon sources and to discriminate between aquatic and terrestrial organic matter⁸.

The CO₂ emissions predicted by Raymond and co-workers' analysis are largest from tropical rivers and lakes in southeast Asia and Amazonia. Because tropical regions are seriously under-represented in global data sets, additional studies of carbon concentrations in the predicted hotspot areas in the tropics are urgently needed. Efforts to constrain data for global emissions of methane — a potent greenhouse gas — from inland waters are also a high priority⁹.

Two other fundamental issues need to be addressed in future work. First, gas transfer along river networks can be dominated by high emission rates at local discontinuities, such as weirs, rapids, waterfalls or turbine releases in hydropower plants⁹. It is therefore

questionable whether a continuous model for gas-transfer velocities based on large-scale geographical parameters, such as that used by Raymond *et al.*, represents the most adequate description of the gas-transfer process.

ANTIBIOTICS

Killing the survivors

Antibiotic-tolerant, dormant variants of otherwise antibiotic-sensitive bacteria underlie many chronic and relapsing infections. A small molecule has been identified that can efficiently eradicate these persister cells. [SEE ARTICLE P.365](#)

KENN GERDES & HANNE INGMER

It is well documented that the incidence of pathogenic antibiotic-resistant bacteria is increasing at an alarming pace. But it is less well known that almost all bacteria, including major pathogens, generate persisters — slow-growing or hibernating cells that are tolerant to multiple antibiotics. Importantly, these variants form at frequencies higher than mutation rates and, consistent with this, seem to be genetically identical to the antibiotic-sensitive organism. Persister cells are a primary source of chronic and relapsing bacterial infections¹ because they are difficult or impossible to eradicate using conventional antibiotics. There is therefore a pressing need to develop treatments that can kill persisters. On page 365

The other issue concerns the heavy modifications that have been made to surface water systems during the past two centuries by channelization and damming. For example, cutting off and draining the wetlands of the lower Danube has reduced the seasonally flooded area of the river corridor by 72%, resulting in an artificial river morphology that cannot be predicted by geographical scaling laws¹⁰. Raymond and colleagues' surprising results call for more specific investigations of how hydraulic constructions in river systems affect global biogeochemical cycles. ■

Bernhard Wehrli is at the *Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, 8092 Zurich, Switzerland, and at the Swiss Federal Institute of Aquatic Science and Technology, Kastanienbaum.*
e-mail: bernhard.wehrli@eawag.ch

1. Raymond, P. A. *et al.* *Nature* **503**, 355–359 (2013).
2. Battin, T. J. *et al.* *Nature Geosci.* **2**, 598–600 (2009).
3. Aufdenkampe, A. K. *et al.* *Front. Ecol. Environ.* **9**, 53–60 (2011).
4. Stumm, W. & Morgan, J. J. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters* 3rd edn (Wiley, 1995).
5. Raymond, P. A. *et al.* *Limnol. Oceanogr. Fluids Environ.* **2**, 41–53 (2012).
6. Downing, J. A. *et al.* *Limnol. Oceanogr.* **51**, 2388–2397 (2006).
7. Zurbrugg, R., Suter, S., Lehmann, M. F., Wehrli, B. & Senn, D. B. *Biogeochemistry* **10**, 23–38 (2013).
8. Feng, X. *et al.* *Proc. Natl. Acad. Sci. USA* **110**, 14168–14173 (2013).
9. Abril, G. *et al.* *Glob. Biogeochem. Cycles* **19**, GB4007 (2005).
10. Tockner, K., Uehlinger, U. & Robinson, C. T. *Rivers of Europe* (Elsevier, 2009).

of this issue, Conlon *et al.*² present evidence that acyldepsipeptides, an emerging class of antibiotic, efficiently kill persisters of certain types of bacterium. Remarkably, when used in conjunction with conventional antibiotics, one of these agents completely eliminated the bacterium *Staphylococcus aureus* growing in culture, and also cured an *S. aureus* infection in mice.

The phenomenon of persistence was discovered almost 70 years ago, but only recently^{3,4} was strong support obtained for the hypothesis that persisters are rare, slow-growing, bacterial cells genetically identical to the rest of the population (typically occurring at a frequency of 1 in 10,000 to 1 in 1 million cells in a rapidly growing population). Phenotypic heterogeneity (in which genetically identical cells exhibit

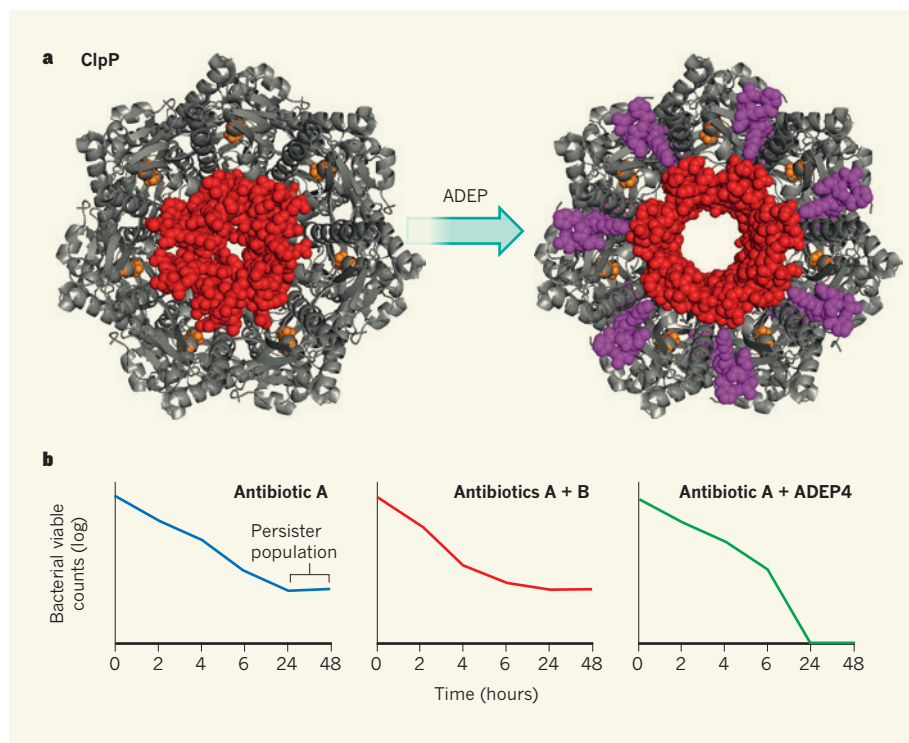


Figure 1 | ADEPs activate ClpP and kill persister bacteria. **a**, Clp is a bacterial protein-degrading enzyme. Its proteolytic subunit ClpP forms a ring-shaped barrel containing a small pore (amino acids lining the pore are shown in red). The pore is normally gated by ClpP-associated ATPase enzymes, which control the entry of protein substrates into the ClpP chamber. Acyldepsipeptide (ADEP) molecules (purple) that bind to the ATPase docking sites on ClpP cause a conformational change that widens the pore, leading to deregulated protein degradation. (Structures provided by Lars Konermann¹³.) **b**, Conlon *et al.*² show that the combination of ADEP4 and a conventional antibiotic kills non-growing persister cells — the small, slow-growing population of cells that persist during treatment with conventional antibiotics (alone or in combination).

different traits) among bacteria can be viewed as a bet-hedging strategy that has evolved to increase survival under environmental insults that could otherwise eradicate the entire population⁵.

The common signalling molecule guanosine tetraphosphate (ppGpp; also called 'magic spot') can render bacterial cells persistent, and is synthesized either by stochastic switching (as in the bet-hedging strategy)⁴ or in response to environmental stress⁶. But because most existing antibiotics target actively growing bacteria, they do not kill these magic-spot-protected persisters.

Acyldepsipeptides (ADEPs) efficiently kill Gram-positive bacteria (a broad bacterial group that includes several human pathogens, including *S. aureus*) at astonishingly low concentrations, and are effective against cells that are resistant to many other antibiotics⁷. Bacterial killing by ADEPs is the result of uncontrolled activation of a protein called ClpP (ref. 7), which is a subunit of the protease enzyme Clp. This enzyme is found in all bacteria and it functions in protein-quality control and in the regulated degradation of specific proteins. The protein-degrading subunit of Clp consists of two heptameric ClpP rings with small central pores through which peptide

chains are threaded into the central proteolytic chamber, after being unfolded by ClpP-associated ATPase enzymes^{8,9}. (The small size of the pore explains why ClpP is largely inactive without these protein-unfolding enzymes.) Structural studies have revealed that ADEPs bind directly to ClpP, independently of the ATPases, and dramatically increase the size of the central pore in ClpP (Fig. 1a). This provides unregulated access for peptides and proteins to the proteolytic chamber^{10,11}, and the resulting increase in protein degradation leads to cell death.

Conlon *et al.* studied non-growing *S. aureus* cells reminiscent of the persisters formed during infection, and found that application of the compound ADEP4 caused the degradation of more than 400 bacterial proteins. This observation led to the straightforward and easily tested hypothesis that ADEP4 would kill persisters, and the authors found that the drug efficiently killed non-growing, stationary-phase bacteria. Remarkably, in combination with a conventional antibiotic (rifampicin, linezolid or ciprofloxacin) that does not kill persisters when administered alone, ADEP4 totally eradicated persisters from a flask culture (Fig. 1b), from a biofilm (a distinct form of aggregated bacteria growing on a surface)

and from an *in vitro* hollow-fibre model used for assessing the efficacy of antibiotics. Even more strikingly, the authors found that ADEP4 eradicated severe, deep-seated *S. aureus* infections in the thighs of mice.

These results raise several key questions. First, will bacteria develop resistance to ADEPs and other compounds that kill bacteria by activating bacterial proteases? Bacteria almost always rapidly become resistant to new antibiotics, so it is to be expected that this will also be the case with ADEPs. But general proteases such as Clp are typically either essential for bacterial survival or are required for bacterial virulence. Therefore, if spontaneous *S. aureus* mutants arise that lack ClpP and are resistant to ADEPs, they will display greatly reduced virulence⁹. Second, will the breakthrough presented by Conlon and colleagues lead to the development of more-effective antibiotics for treating relapsing and chronic infections? We believe that this is probable. Most antibiotics that actively kill bacteria do so by corrupting a cellular target that is particularly active during bacterial growth, whereas ADEPs activate their cellular target whether the bacteria are growing or not. Unsurprisingly, the research group is now testing a second class of antibiotic that activates ClpP (ref. 12). This growing body of results generates hope that antibiotics for the treatment of persistent infections will be available in the future. ■

Kenn Gerdes is at the Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4AX, UK, and in the Department of Biology, University of Copenhagen, Denmark. **Hanne Ingmer** is in the Department of Veterinary Disease Biology, University of Copenhagen, 1870 Frederiksberg, Denmark.
e-mails: kenn.gerdes@newcastle.ac.uk; hi@sund.ku.dk

- Lewis, K. *Annu. Rev. Microbiol.* **64**, 357–372 (2010).
- Conlon, B. P. *et al. Nature* **503**, 365–370 (2013).
- Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. *Science* **305**, 1622–1625 (2004).
- Maisonneuve, E., Castro-Camargo, M. & Gerdes, K. *Cell* **154**, 1140–1150 (2013).
- Veening, J. W., Smits, W. K. & Kuipers, O. P. *Annu. Rev. Microbiol.* **62**, 193–210 (2008).
- Dalebroux, Z. D. & Swanson, M. S. *Nature Rev. Microbiol.* **10**, 203–212 (2012).
- Brötz-Oesterhelt, H. *et al. Nature Med.* **11**, 1082–1087 (2005).
- Baker, T. A. & Sauer, R. T. *Biochim. Biophys. Acta* **1823**, 15–28 (2012).
- Frees, D., Savijoki, K., Varmanen, P. & Ingmer, H. *Mol. Microbiol.* **63**, 1285–1295 (2007).
- Li, D. H. *et al. Chem. Biol.* **17**, 959–969 (2010).
- Lee, B. G. *et al. Nature Struct. Mol. Biol.* **17**, 471–478 (2010).
- Leung, E. *et al. Chem. Biol.* **18**, 1167–1178 (2011).
- Sowole, M. A., Alexopoulos, J. A., Cheng, Y. Q., Ortega, J. & Konermann, L. J. *Mol. Biol.* **425**, 4508–4519 (2013).

This article was published online on 13 November 2013.

CLIMATE SCIENCE

The challenge of hot drought

An analysis of North American drought variability over the past millennium shows that it is not unusual for widespread drought to persist for years, prompting fresh thinking about our ability to deal with such climate conditions.

JONATHAN T. OVERPECK

Drought is heating up around the warming world. Particularly hot drought has cost more than US\$40 billion and claimed 218 human lives since 2010 in the United States alone¹. These hot and dry conditions have also contributed to unusually widespread and devastating wildfires¹, fuelled by wide expanses of weakened and dead trees that were unable to deal with heat stress and subsequent insect attack². Yet, to get a real sense of how this recent change in drought severity might shape the future, one has to look to the past. An analysis of regional and pan-continental North American drought over the past 1,000 years, reported by Cook *et al.*³ in the *Journal of Climate*, makes it clear that recent droughts, as costly as they have been, are only a taste of what might lie ahead, independently of any big climate change.

Drought conditions, including the two most severe categories — extreme and exceptional —

covered more than half of the continental United States in 2012⁴. This drought affected several regions of North America (Fig. 1), earning it the distinction of being a pan-continental drought rather than the more common regional drought³. Cook *et al.* tap a continental array of 1,000-year drought records based on tree rings to show how the 2012 pan-continental drought pattern has occurred in 12% of years since the tenth century. More importantly, the authors' study highlights how no major US region is immune to such drought, and that we understand quite a lot about how sea surface temperatures drive the differing patterns of drought.

Cook and colleagues' most relevant lesson for the future, however, may be that the one-year pan-continental drought of 2012 was but a glimpse of drought compared with the multi-decadal pan-continental megadroughts that were most common during the twelfth and thirteenth centuries. The complexity of these megadroughts still defies complete

explanation and yet it implies that unusually persistent anomalies of sea surface temperature can combine with amplifying changes in vegetation and soil to drive droughts that — if they happened today — would outstrip many of our institutional capacities to deal with such aridity. For example, another tree-ring study⁵ highlighted a 50-year drought, with only one normal year of precipitation, in the headwaters of both the Colorado River and the Rio Grande during Roman times. It is hard to imagine how such a drought would play out today, but it would surely prove a much greater challenge to regional water resources and forests than any drought of the past 120 years.

Tree-ring records are just one invaluable source of palaeoclimatic information. Proxy climate records from lake sediments and cave formations also help to show how drought has varied over timescales that are too long to be understood from the short record provided by thermometers and rain gauges alone. Moreover, palaeoclimatic data provide long records of climate variability against which state-of-the-art models can be tested. For example, the climate models used in the ongoing Fifth Assessment of the Intergovernmental Panel on Climate Change (IPCC) seem to underestimate the strong multi-decadal drought variability that is evident in the multi-proxy palaeoclimatic record⁶. This implies that, although well validated in many ways, these models underestimate the risk of future multi-decadal megadroughts of the type that plagued medieval and earlier times in the southwest United States. There are many potential reasons for this shortcoming, including perhaps inadequate representation of tropical Pacific variability in the models^{6,7}.

Cook and colleagues end their new work with a warning that global warming has the potential to increase the severity and extent of future droughts. This seems clear in some regions, such as the southwest United States, for which researchers have coined the term "global-change-type drought"⁸, which might more appropriately be called global-warming-type drought. Warming seems already to be altering the duration and frequency of drought in some regions of the globe, a trend that will probably become clearer as global warming proceeds⁹. In addition, warming is likely to reduce flows in snow-fed rivers such as those of the western United States^{10,11}, and will intensify vegetation stress during drought¹².

The bottom line on drought seems evident. Cook *et al.* highlight the rich diversity of droughts that can occur naturally. Droughts can be short, and they can persist for decades. Moreover, they can be intensely regional or be pan-continental. Without doubt, the situation calls for the public, policy-makers and all types of resource managers to consider 'no-regrets options' (activities that yield benefits no matter what) for dealing with long and potentially extensive drought that will inevitably happen

MAP: BENJAMIN COOK; DATA SOURCE: CLIMATIC RESEARCH UNIT/UNIV. EAST ANGLIA

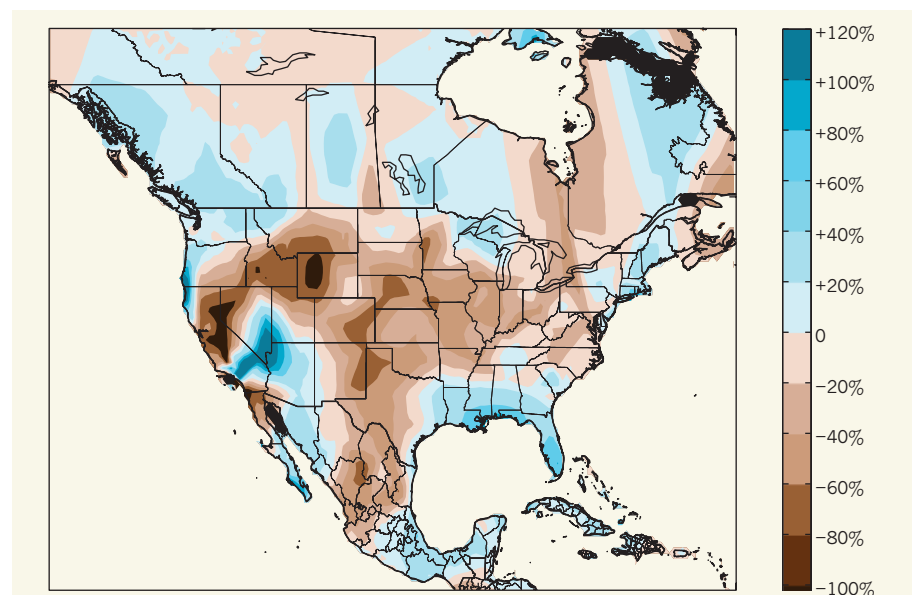


Figure 1 | The 2012 North American drought. The map shows the per cent precipitation anomaly for June–August 2012, relative to the mean for 1961–90, and illustrates (in brown colours) the widespread nature of the 2012 drought. Although this three-month period was wetter than normal in parts of the southwest United States, this region has also been in drought more often than not since 1999. Palaeoclimatic records³ indicate that droughts experienced over the past 100 years, including the costly 2012 drought, have been modest compared with the often much longer and equally widespread droughts of the past millennium.

again in the future. Choices include saving groundwater for when it is really needed; injecting extra water into the ground during wet periods for storage; making water use more efficient; perfecting the inexpensive reuse of water; and maintaining water use in activities (such as farming) whose users can sell their water when less-flexible users (including urban populations) need help in dealing with extended drought. These strategies might prove more feasible than massive efforts to transfer water between regions, especially given that, as Cook *et al.* show, many regions can be hit by drought simultaneously. At the same time, work is urgently needed to understand all the ways in which global climate change will exacerbate the types of drought that have occurred before.

Could we thrive in the face of a pan-continental megadrought? Probably yes, but

only if we develop the appropriate buffering strategies in advance of the drought, and if we think more seriously about reducing emissions of greenhouse gases to the atmosphere as a way of keeping future droughts from becoming so hot and dry that they are beyond our buffering capacity. ■

Jonathan T. Overpeck is at the Institute of the Environment and in the Departments of Geosciences and of Atmospheric Sciences, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: jto@email.arizona.edu

1. NOAA National Climatic Data Center. *Billion-Dollar Weather/Climate Disasters* www.ncdc.noaa.gov/billions/events
2. Williams, A. P. *et al. Proc. Natl Acad. Sci. USA* **107**, 21289–21294 (2010).

3. Cook, B. I., Smerdon, J. E., Seager, R. & Cook, E. R. *J. Clim.* <http://dx.doi.org/10.1175/JCLI-D-13-00100.1> (2013).
4. United States Drought Monitor <http://droughtmonitor.unl.edu>
5. Routson, C. C., Woodhouse, C. A. & Overpeck, J. T. *Geophys. Res. Lett.* **38**, L22703 (2011).
6. Ault, T. R. *et al. J. Clim.* **26**, 5863–5878 (2013).
7. Ault, T. R., Cole, J. E. & George, S. S. *Geophys. Res. Lett.* **39**, L21705 (2012).
8. Breshears, D. D. *et al. Proc. Natl Acad. Sci. USA* **102**, 15144–15148 (2005).
9. Stocker, T. F. *et al. (eds) Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers* (IPCC, 2013).
10. Das, T., Pierce, D. W., Cayan, D. R., Vano, J. A. & Lettenmaier, D. P. *Geophys. Res. Lett.* **38**, L23403 (2011).
11. Vano, J. A., Das, T. & Lettenmaier, D. P. *J. Hydrometeorol.* **13**, 932–949 (2012).
12. Williams, A. P. *et al. Nature Clim. Change* **3**, 292–297 (2013).

HUMAN EVOLUTION

Group size determines cultural complexity

Many animals use culture, the ability to learn from others, but only humans create complex culture. A laboratory experiment tests which characteristics of our social networks give us this capacity. SEE LETTER P.389

PETER RICHESON

Isaac Newton famously said that he saw further by standing on the shoulders of giants. A more apt image for most human culture is that we see further because we stand on the shoulders of a vast pyramid of mini-Newtons. Only a few people have invented even one word of the language they speak, for example, yet a native speaker of English knows tens of thousands of words. As early as the Stone Age, people spoke complex languages, interacted in diverse social systems and built exquisite and functional tools. So how do we create the wonderfully diverse cultural systems that sustain us in almost every terrestrial habitat in the world? Studies of cultural evolution point to two factors — accurate imitation¹ and large social networks². Mathematical modelling suggests³ that these two properties will support the fast, cumulative evolution of cultural systems. On page 389 of this issue, Derex *et al.*⁴ present results from a laboratory experiment that support the role of network size (Fig. 1).

Accurate imitation allows humans, but not chimpanzees, to learn complex skills and ideas from others — much more complex ones than they can learn for themselves. Large social networks allow human learners to tap the knowledge of mentors skilled in any

cultural domain, thereby rapidly spreading the best ideas throughout a society. Studies to test the effect of network size on cultural evolution have mainly used observations of small, isolated populations compared with larger neighbouring groups⁵. But such natural experiments are controversial: not all studies

find the effect, perhaps because other factors also influence cultural complexity. Therefore, Derex *et al.* turned to the laboratory to investigate the issue.

Theory suggests² that if a too-small group attempts to make a too-complex tool, over time the tool will become simplified: small groups will often lack a tool-maker of sufficient skill to make the complex version of the tool and a simpler form will evolve. To study the effects of varying task complexity and the number of members in groups of learners, Derex *et al.* asked participants to draw either a stylized arrowhead or a fishing net on a computer screen. These designs were then used to earn the participants money from simulated hunting or fishing expeditions. The monetary yield of an arrowhead was a simple function of its shape, whereas that for nets was a complex function of net shape, the size of cord used in different parts of the net and



Figure 1 | Net gain. Derex *et al.*⁴ show that interacting in large groups helps people to maintain the ability to perform complex tasks, such as building nets.

GUNTHER GRÄFENHAIN/ACORNERS IMAGES

the knots used to hold the cords together. The yield from a well-constructed net was considerably more than could be earned from an arrowhead, so participants were motivated to construct nets.

The participants were assigned to groups of 2, 4, 8 or 16. They received initial video demonstrations in how to make both tools and then had 15 trials to make their own — one tool per turn. At the end of each trial, participants could see the yield of each of the other people in their group, and by clicking on those scores, could see the step-by-step procedure by which the corresponding object had been made.

The authors' findings support the hypothesis that group size plays an important part in cultural evolution: the probability of a group maintaining the ability to construct the complex tool (the net) over the course of the experiment, the probability of maintaining the ability to construct both tools, and the quality of both tools all increased as a function of group size. Most participant attempts to copy a demonstration for making the fishing net resulted in nets worse than the original. Nevertheless, in large groups, the best nets were often better than the demonstration, and this drove the maintenance of net quality in those groups, as predicted by theory. By contrast, net quality deteriorated substantially in smaller groups. The quality of the arrowheads improved considerably over the course of the trials in the larger groups and was more or less maintained in smaller groups.

A noteworthy wrinkle in the findings is that the performance in groups of 8 and 16 participants hardly differed, perhaps because the extra information in groups of 16 was as distracting as it was helpful. Furthermore, participants were under time pressure in observing others' procedures and making their new tools.

Laboratory experiments have the obvious problem of drastically compressing the time-scale of social learning and cultural evolution, and the size of populations. But despite the difficulties of capturing culture in the laboratory, the need to do so is overwhelming. Cultural transmission is much messier than genetic transmission. The extended duration of enculturation, and the involvement of ill-defined and interacting influences, make studying cultural transmission in almost all natural populations difficult compared with studying the discrete events and the one or two parents involved in genetic reproduction. In addition, learners' own preferences also influence what is transmitted, and this situation is without parallel in biological reproduction. Controlled experiments are the only way to understand many of these processes and, as in so many fields, the problem of laboratory artefacts must be considered part of the price.

Although proposals to conduct such experiments go back a long way⁶, and some older attempts produced interesting results⁷, culture

researchers are only at the beginning of their experimental project — they are essentially a century behind geneticists working on a similar project. The field of cultural evolution has grown up at the intersection of disparate disciplines and initial progress was slow. Evolutionary biologists and economists furnished the formal theory; anthropologists, sociologists and historians contributed their interest in culture; and social and developmental psychologists brought a focus on individuals and methods for studying how individuals interact with their groups. But only recently have experiments like those of Derex and colleagues been appreciated by a broad audience.

Science itself is a cultural evolutionary phenomenon, and understanding it as such is an important project in itself. The polymath psychologist and pioneering contributor to the study of cultural evolution, Donald T. Campbell, proposed an applied cultural-evolution project designed to improve scientific practice⁸. Recently, an article⁹ in *The Economist* was featured on the magazine's cover as 'How science goes wrong'. Campbell's rarely discussed

idea seems worth pursuing as part of our continuing studies of cultural evolution. ■

Peter Richerson is in the Department of Environmental Science and Policy, University of California, Davis, Davis, California 95616, USA, and at the School of Archaeology, University College London, UK.
e-mail: pjricherson@ucdavis.edu

1. Whiten, A., McGuigan, N., Marshall-Pescini, S. & Hopper, L. M. *Phil. Trans. R. Soc. B* **364**, 2417–2428 (2009).
2. Henrich, J. *Am. Antiq.* **69**, 197–214 (2004).
3. Cavalli-Sforza, L. L. & Feldman, M. W. *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton Univ. Press, 1981).
4. Derex, M., Beugin, M.-P., Godelle, B. & Raymond, M. *Nature* **503**, 389–391 (2013).
5. Kline, M. A. & Boyd, R. *Proc. R. Soc. B* **277**, 2559–2564 (2010).
6. Gerard, R. W., Kluckhohn, C. & Rapoport, A. *Behav. Sci.* **1**, 6–34 (1956).
7. Insko, C. A. et al. *J. Pers. Social Psychol.* **44**, 977–999 (1983).
8. Campbell, D. T. *Sci. Technol. Hum. Values* **10**, 38–48 (1985).
9. *The Economist* go.nature.com/dstij3 (2013).

This article was published online on 13 November 2013.

HIV

Slipping under the radar

HIV avoids triggering the cell receptors that initiate the host's innate immune responses. It seems that the virus achieves this evasion by using its protein coat to hide its nucleic acids until they are beyond detection. [SEE LETTER P.402](#)

STEPHEN P. GOFF

Many cells use an elaborate warning system to detect and respond to viral infection. A variety of viral proteins and nucleic-acid structures, called pathogen-associated molecular patterns, serve as warning flags for infection (for a review, see ref. 1). Detection of one of these patterns can trigger a cell to produce interferons and other cytokines — cell-signalling molecules that induce the expression of antiviral genes in neighbouring cells, thereby establishing in them a protective antiviral state. Unlike many other viruses, retroviruses such as HIV-1 tend to be poor inducers of interferons, although they are highly sensitive to inhibition by interferon-stimulated gene products if these genes are artificially induced. How these viruses evade detection has been unclear, but in this issue, Rasaiyaah et al.² (page 402) provide intriguing evidence that HIV-1 slips under the radar of the surveillance system by using its viral capsid — the protein shell surrounding the viral genome — to hide its nucleic acids during the

early stages of infection. The authors show that disrupting the proper capsid structure or the timing of the infection can expose this hidden viral DNA to detection.

The events occurring immediately after retroviral infection are among the least-understood parts of the viral life cycle. Infection is known to result in the delivery of a viral core particle into a cell's cytoplasm, and a progressive 'uncoating' of this particle (although exactly what is meant by this term is not clear) accompanies subsequent stages of infection: the reverse transcription of the virus's RNA genome into DNA; the trafficking of the core particle towards the nucleus; its import into the nucleus; and, ultimately, the integration of viral DNA into the host chromosomal DNA. At least some of the HIV-1 capsid proteins are present in all of these steps, even at the time of nuclear DNA integration. The capsid interacts with several host proteins during infection, including CPSF6 (ref. 3), Nup153 (ref. 4), cyclophilin A (ref. 5), and other proteins with cyclophilin domains, such as Nup358 (ref. 6).

These interactions can determine how the

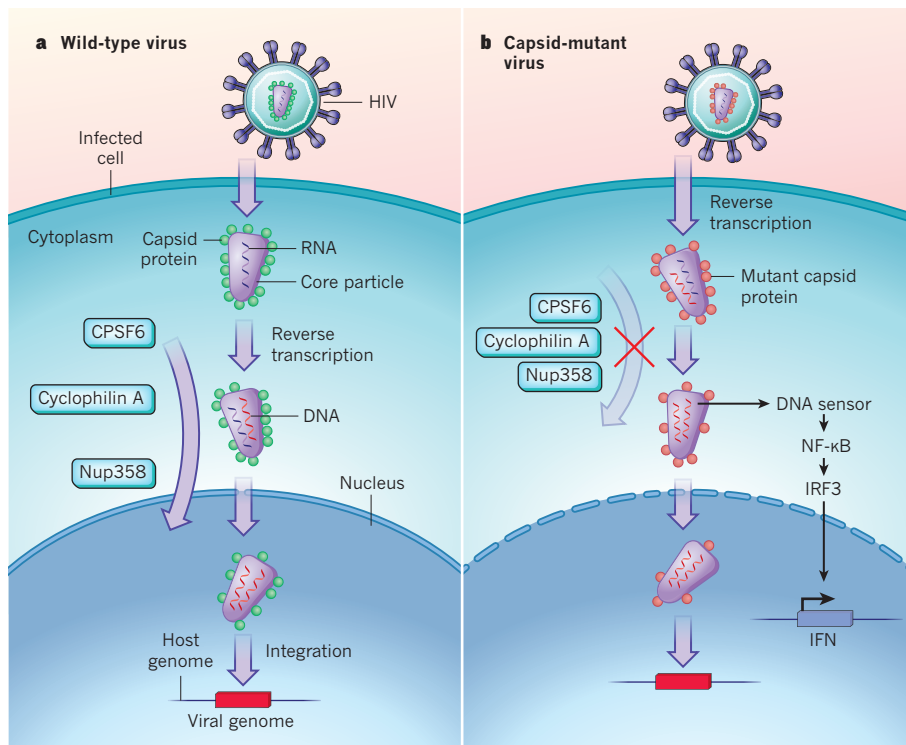


Figure 1 | Retroviral evasion of immune detection. **a**, During infection with wild-type HIV, the viral core particle enters the cytoplasm of the infected cell, where its viral RNA is reverse transcribed into DNA. The particle then moves to the nucleus, and the viral DNA is integrated into the host genome. During these stages of infection, the protein capsid surrounding the viral particle interacts with host-cell proteins, including CPSF6, cyclophilin A and Nup358. Wild-type HIV does not trigger the production of the antiviral cell-signalling molecule IFN. **b**, By contrast, Rasaiyaah *et al.*² show that IFN production is triggered in response to infection with HIV viruses carrying mutations in capsid proteins that block the proteins' interactions with host proteins. These mutant capsids may traffic differently within the cell, and reverse transcription may occur prematurely, allowing recognition of viral DNA by host-cell sensors that induce a signalling pathway (involving the transcription factors NF-κB and IRF3) leading to IFN production. The altered host–capsid protein interactions may also change the mode of virus entry into the nucleus, allowing access only while the nuclear membrane is disrupted during cell division.

virus enters the nucleus, whether it requires cell division and the resulting disruption of the nuclear membrane, and whether it makes use of specific nuclear-pore proteins⁶. There are known capsid mutations that affect these interactions and they have profound effects on the timing and route of nuclear entry (Fig. 1). Rasaiyaah *et al.* now report the surprising finding that viruses with either of two such capsid mutations induce high levels of type 1 interferon (IFN) in infected cells, implying that the capsid normally has a crucial role in preventing virus DNA detection. The results have intriguing implications for our understanding of the functions of the capsid and of capsid-interacting proteins.

The authors show that wild-type HIV-1 does not trigger the pattern-recognition sensors of the host innate immune system, and that it grows well in cells called human monocyte-derived macrophages. But they show that mutant viruses with either of two single-nucleotide mutations in the gene encoding the capsid protein — mutant N74D, which fails to interact with CPSF6 (refs 3, 7), and mutant P90A,

which does not interact with cyclophilin A or Nup358 — do induce type 1 IFN and cannot replicate in these macrophages. IFN-stimulated genes were induced following infection with the mutant viruses, although not by the wild-type virus. But IFN induction and blocking of viral replication did occur in response to infections with the wild-type virus if CPSF6 was depleted or if the cyclophilins were inhibited. In all of these settings, it seems that IFN induction is the major cause of the block to viral replication, because Rasaiyaah *et al.* show that preventing IFN signalling by using antibodies to the IFN receptor restored normal replication of the mutant viruses.

Further experiments suggested that the key trigger of IFN production is viral DNA formed by reverse transcription soon after virus entry into the cell. Viruses carrying both the activating capsid mutations and also mutations affecting the reverse transcriptase enzyme did not induce IFN, whereas IFN induction did occur with viruses that had mutations in the integrase enzyme (responsible for the integration of viral DNA into the host genome). Thus, it seems that some form of viral DNA, but not

integrated DNA, is the trigger. The sensing pathway for the P90A mutant, but perhaps not the N74D mutant, probably involves the production of cyclic GMP–AMP, a newly described signalling intermediate formed in response to cytosolic viral DNA⁸.

The role of CPSF6 in viral uncoating and sensing is particularly complicated. It seems to have either positive or inhibitory effects on virus replication in different settings. Over-expression of a fragment of CPSF6 in the cytoplasm can cause binding of the capsid and inhibition of virus replication, perhaps by inhibiting uncoating³. But without CPSF6 binding (as occurs with the N74D mutant virus), or when CPSF6 is depleted, the virus triggers IFN production, perhaps because of premature DNA synthesis. The functions of this host protein are further complicated by its interaction with TNPO3, a transporter protein that may be able to control its intracellular localization. These proteins may be involved in the import of viral DNA into the nucleus. The only simple punchline may be that CPSF6 plays an important part in the timing of reverse transcription and/or uncoating, and thus in evading virus detection. Cyclophilin A may have a similar, although not identical, role.

Rasaiyaah and colleagues' findings provide some big surprises about how retroviruses evade innate immunity. They also raise questions about the detection of the viral DNA. When and where in the cell does the crucial part of reverse transcription occur? Exactly what is the DNA-containing structure that is sensed by the detection system, and what is the proximal detector of the incoming viral DNA? What cell types are capable of detecting infection? (Some data indicate that not all cells respond equally to infection by HIV-1 capsid mutants.) And, now that we better understand how the wild-type virus has evolved to escape detection, can we somehow reinstruct the system to work around these escape mechanisms, successfully detect the virus, induce IFN production and block replication? More information is sure to be forthcoming soon. ■

Stephen P. Goff is in the Departments of Microbiology & Immunology and Biochemistry & Molecular Biophysics, Columbia University, New York, New York 10032, USA.
e-mail: spg1@columbia.edu

1. Melchjorsen, J. *Viruses* **5**, 470–527 (2013).
2. Rasaiyaah, J. *et al.* *Nature* **503**, 402–405 (2013).
3. Lee, K. *et al.* *Cell Host Microbe* **7**, 221–233 (2010).
4. Matreyek, K. A. & Engelman, A. J. *Virology* **85**, 7818–7827 (2011).
5. Luban, J., Bossolt, K. L., Franke, E. K., Kalpana, G. V. & Goff, S. P. *Cell* **73**, 1067–1078 (1993).
6. Schaller, T. *et al.* *PLoS Pathog.* **7**, e1002439 (2011).
7. Ambrose, J. *et al.* *J. Virol.* **86**, 4708–4714 (2012).
8. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. *Science* **339**, 786–791 (2013).

This article was published online on 6 November 2013.

Global carbon dioxide emissions from inland waters

Peter A. Raymond¹, Jens Hartmann^{2*}, Ronny Lauerwald^{2,3*}, Sebastian Sobek^{4*}, Cory McDonald⁵, Mark Hoover¹, David Butman^{1,6}, Robert Striegl⁶, Emilio Mayorga⁷, Christoph Humborg⁸, Pirkko Kortelainen⁹, Hans Dürr¹⁰, Michel Meybeck¹¹, Philippe Ciais¹² & Peter Guth¹³

Carbon dioxide (CO₂) transfer from inland waters to the atmosphere, known as CO₂ evasion, is a component of the global carbon cycle. Global estimates of CO₂ evasion have been hampered, however, by the lack of a framework for estimating the inland water surface area and gas transfer velocity and by the absence of a global CO₂ database. Here we report regional variations in global inland water surface area, dissolved CO₂ and gas transfer velocity. We obtain global CO₂ evasion rates of $1.8^{+0.25}_{-0.25}$ petagrams of carbon (Pg C) per year from streams and rivers and $0.32^{+0.52}_{-0.26}$ Pg C yr⁻¹ from lakes and reservoirs, where the upper and lower limits are respectively the 5th and 95th confidence interval percentiles. The resulting global evasion rate of 2.1 Pg C yr^{-1} is higher than previous estimates owing to a larger stream and river evasion rate. Our analysis predicts global hotspots in stream and river evasion, with about 70 per cent of the flux occurring over just 20 per cent of the land surface. The source of inland water CO₂ is still not known with certainty and new studies are needed to research the mechanisms controlling CO₂ evasion globally.

Quantifying the Earth's global carbon cycle is essential for a sustainable future because CO₂ has an active role in the Earth's energy budget. Natural ecosystems are important to this accounting because they exchange large amounts of CO₂ with the atmosphere and currently offset $\sim 4 \text{ Pg C yr}^{-1}$ of anthropogenic emissions¹. Until now, estimates of the global exchange of CO₂ between inland waters and the atmosphere have not been made using comprehensive, spatially resolved efforts. It was shown definitively 30 years ago that CO₂ in inland waters calculated from alkalinity and pH were substantially higher than atmospheric values². Early direct measurements of large rivers and arctic inland waters also demonstrated supersaturation^{3–6}. The first regional estimate of inland water degassing, which was for the Amazon, was not reported until 2002⁷. That study estimated the release of $\sim 0.5 \text{ Pg C yr}^{-1}$ (ref. 7) from streams, rivers and wetlands of this region alone, and was revised upwards to account for a large degree of CO₂ supersaturation in small headwater streams⁸. Recently, the total CO₂ emitted from the contiguous United States streams and rivers was estimated at $\sim 0.1 \text{ Pg C yr}^{-1}$, extrapolated to 0.5 Pg C yr^{-1} for temperate rivers between latitudes 25° and 50° north⁹.

There are few global estimates of inland waters CO₂ evasion^{10–13}. These studies still place the efflux at only $\sim 1 \text{ Pg C yr}^{-1}$ (refs 10–13), despite the high fluxes estimated for temperate rivers and the Amazon. To date, global exchange calculations are simple in nature and prone to uncertainties in all three factors that determine inland water CO₂ evasion: the amount of CO₂ in water; the global surface area of streams, rivers, lakes and reservoirs; and the gas transfer velocity (k , a parameter that relates to the physics that determines the rate of gas exchange). Recently, studies have revisited the scaling of lake and reservoir surface area, using new geospatial data sets^{14–16} that we adapted to produce spatially explicit

global maps of lake and reservoir surface area divided by size classes. Other studies have also probed the controls on and the quantities of lake-dissolved CO₂ at the large catchment scale^{17–20}, and have improved our knowledge of the drivers of the gas transfer velocity in lake and reservoir systems^{21,22}, which we synthesized here for our global estimate.

Studies in rivers and streams have also progressed. Regional studies have attempted a more systematic estimation of stream and river evasion for Sweden, the United States and the Yukon River basin^{9,19,23}. This approach entails using stream scaling laws and high-resolution remote-sensing information that exists for these regions. Although similar high-resolution maps are not available globally for streams and rivers, we provide a new spatially resolved global stream surface area and gas transfer velocity using coarser global data sets that have recently been developed²⁴, combined with river scaling laws^{25,26}, discharge estimates for global drainage basins²⁷ and new knowledge of the controls on the gas transfer velocity for streams and rivers^{28,29}.

We have combined these new approaches for estimating the global inland water surface area and gas transfer velocity with a new global data set of calculated values of the CO₂ partial pressure, p_{CO_2} (based on the GloRiCh database³⁰), to provide spatial maps of inland water CO₂ evasion along with uncertainty intervals. We perform our scaling using the COSCAT (coastal segmentation and related catchment) drainage network segmentation framework³¹, which lends itself to drainage basin analysis and allows for the spatial representation of this exchange.

Inland water surface area

We find a strong positive correlation between stream and river surface area and precipitation, and a weaker negative relationship between surface area and temperature (Supplementary Fig. 4). The robust

¹Yale School of Forestry and Environmental Studies, 195 Prospect Street, New Haven, Connecticut 06511, USA. ²Institute for Geology, KlimaCampus, Universität Hamburg, D-20146 Hamburg, Germany. ³Department of Earth and Environmental Sciences, Université Libre de Bruxelles, B-1050 Bruxelles, Belgium. ⁴Department of Ecology and Genetics, Limnology, Uppsala University, SE-75236 Uppsala, Sweden. ⁵Wisconsin Department of Natural Resources, Madison, Wisconsin 53716, USA. ⁶US Geological Survey, National Research Program, Boulder, Colorado 80303, USA. ⁷Applied Physics Lab, University of Washington, Seattle, Washington 98105, USA. ⁸Department of Applied Environmental Science, Stockholm University, S-10691 Stockholm, Sweden. ⁹Finnish Environment Institute, PO Box 140, FI-00251 Helsinki, Finland. ¹⁰Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. ¹¹Université Pierre et Marie Curie (Paris VI), Unité Mixte de Recherche CNRS-UPMC Sisyphe, F-75252 Paris 05, France. ¹²LSCE IPSL, UMR8212, F-91191 Gif-sur-Yvette, France. ¹³Department of Oceanography, US Naval Academy, 572C Holloway Road, Annapolis, Maryland 21402, USA.

*These authors contributed equally to this work.

relationship between stream area and precipitation is driven mostly by a strong positive correlation between stream width per Strahler stream order and precipitation, and efforts that use a global average stream width for all streams and rivers will therefore not capture higher surface area of streams and rivers in wetter regions of the globe. Globally we predict a 0.07% increase in the fraction of stream area for a 10-cm increase in precipitation and a 0.02% decrease with an increase in temperature by 1 °C (Supplementary Information). These correlations, which have also been demonstrated with satellite measurements³², are important to global change studies because they reveal a potential link between water cycle changes and inland water surface area.

We first calculate a global stream and river surface area of 624,000 km² (average of 487,000 and 761,000 km², estimated using two different hydraulic equations; Supplementary Information), or 0.47% of the Earth's land surface (Antarctica is excluded from this analysis). The estimate of 624,000 km² does not include ephemeral and intermittent stream fraction periods (Supplementary Information), which removed ~84,000 km² of stream surface area from contributing to gas exchange. This is towards the upper limit of a recent estimate of 485,000–662,000 km² (ref. 33). However, the latter study may not have captured first-order streams, which are included here (Supplementary Information). Previous studies also did not account for spatial variability in width and therefore possibly underestimated the contribution of surface area from wet regions of the globe. Our analysis predicts a large (~15%) contribution to total stream and river surface area from small streams (Supplementary Table 1). We also corrected for the amount of frozen streams with little gas exchange (the effective surface area; see Supplementary Information), further reducing our estimate down to 536,000 km² (Supplementary Information). Using this effective surface area weakens the strength of the negative correlation between temperature and stream surface area. High surface area is estimated in areas of the tropics and temperate regions of the globe (Fig. 1).

We estimate a global lake and reservoir surface area of 3,000,000 km², or 2.2% of the Earth's surface, of which 91.3% is lakes and 8.7% is reservoirs. Our estimate was arrived at using a combination of empirical data for large lakes with statistical models based on regional inventories of smaller lakes (Supplementary Table 4). These estimates of surface area are lower than a recent estimate³⁴ but are proximate to others³⁵. Our lake surface area is lower than some recent estimates because we estimate a smaller contribution from small lakes (Supplementary Table 4) as a result of recent work demonstrating that the size distribution of small lakes is independent of that of large lakes¹⁶. Combining lakes and reservoirs with streams and rivers provides a total surface area of inland waters of 3,620,000 km². High coverage of lakes can be found in previously glaciated landscapes of temperate and arctic regions, and in mountain regions, where glacial movements and tectonic activity have created a multitude of depressions (Fig. 2). It should be noted that the estimate of surface area does not include wetlands. We believe wetlands are functionally different than inland waters because a canopy of vegetation can alter the direction of atmospheric CO₂ exchange.

Inland water CO₂

Inland waters are generally supersaturated with CO₂ with respect to water in equilibrium with the atmosphere. Of the 6,708 stream and river sampling locations for which at least one p_{CO_2} value was calculated, 95% had a median p_{CO_2} greater than atmospheric values (Supplementary Information). The average of these median values was ~2,300 μatm , which increases to an average p_{CO_2} of ~3,100 μatm after discounting for potential biases in the calculation and normalizing interpolated p_{CO_2} from each region to stream area (Supplementary Information). It is important to note that we were not able to assign p_{CO_2} by stream order for this study. An average of 3,100 μatm is within the range of ~1,300–4,300 μatm reported in previous regional or global studies^{7,10,28,36}. The concentration of CO₂ in water was not found to be strongly related to climatic or landscape variables (Supplementary Information), which is consistent with a recent study for North America³⁰,

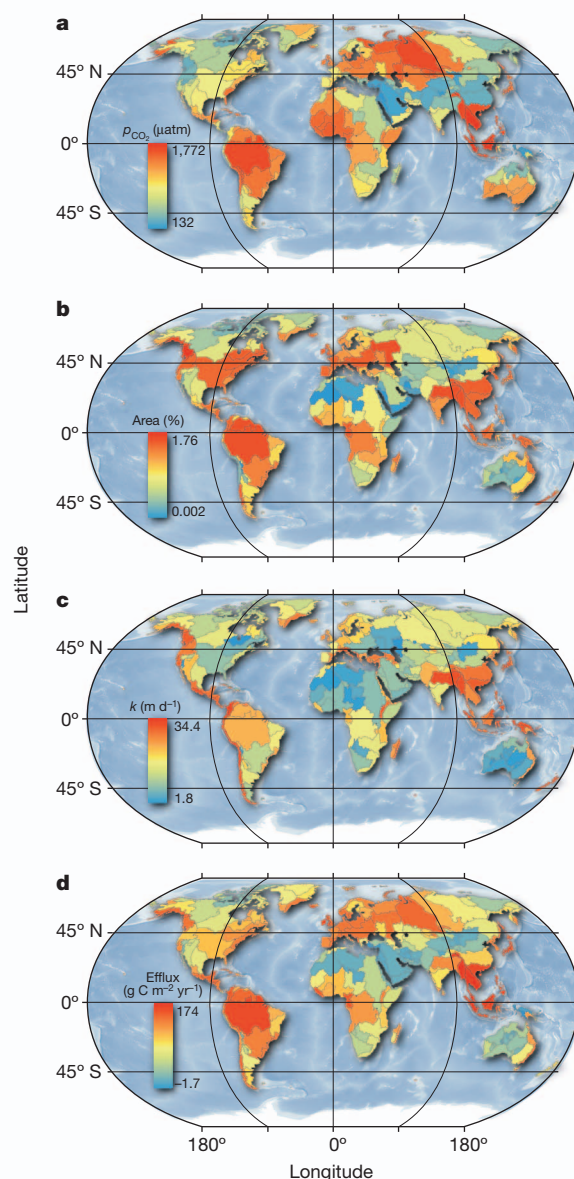


Figure 1 | Maps of stream and river gas exchange parameters. **a**, p_{CO_2} ; **b**, effective surface area; **c**, stream gas transfer velocity; **d**, CO₂ efflux (area normalization is with respect to the area of each COSCAT region).

which showed strong correlations with alkalinity and pH, but weaker correlations between climatic variables and CO₂.

We assembled 20,632 p_{CO_2} observations from 7,939 lakes and reservoirs that were also generally supersaturated. Three groups of lakes could be distinguished on the basis of p_{CO_2} : non-tropical freshwater lakes, tropical lakes and saline lakes. Reservoirs were treated as similar to natural lakes because their p_{CO_2} values have been shown to be elevated only during the initial ~15 yr after impoundment^{37,38}. Non-tropical freshwater lakes had a median p_{CO_2} of 1,120 μatm and a mean of 1,410 μatm (Supplementary Information). Tropical and saline lakes were higher and lower in p_{CO_2} , respectively (Supplementary Information), although these lakes were not well represented in the data set (1.5% and 0.8%, respectively). Also, the respective median values, 1,910 and 270 μatm , were significantly different than the mean values, 4,390 and 1,190 μatm , for tropical and saline lakes. We therefore used the median values to upscale to lakes in tropical and endorheic regions, owing to the potential for overestimation when calculating CO₂ from alkalinity and pH, and to avoid any bias from a few very high p_{CO_2} values (Supplementary Information). In non-tropical freshwater lakes,

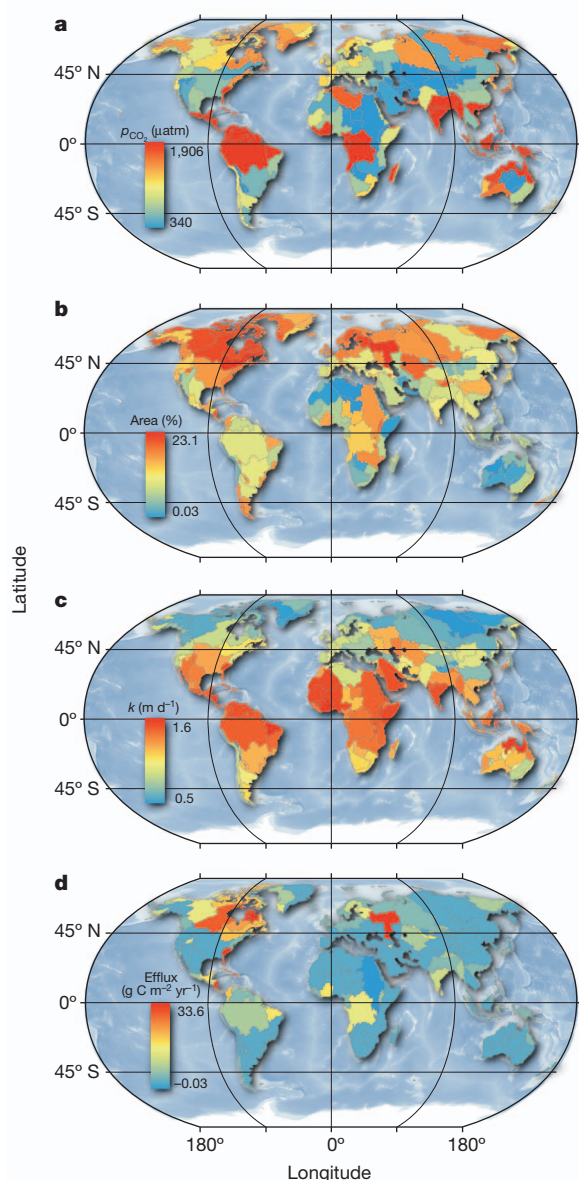


Figure 2 | Maps of lake and reservoir gas exchange parameters. **a**, p_{CO_2} ; **b**, effective surface area; **c**, lake gas transfer velocity; **d**, CO_2 efflux (area normalization is with respect to the area of each COSCAT region).

CO_2 was positively correlated with the concentration of total organic carbon (TOC) and negatively correlated with lake size (Supplementary Information), and these correlations were used to extrapolate lake p_{CO_2} for non-tropical exorheic COSCAT regions of the globe. Globally, dissolved p_{CO_2} normalized to lake area was $\sim 800 \mu\text{atm}$. Lake p_{CO_2} is highest in the humid tropics and also in some boreal regions, owing to high TOC concentrations (Fig. 2).

Inland water gas transfer velocity

The global average gas transfer velocity of 5.7 m d^{-1} (average of 5.0 and 6.3 m d^{-1} , estimated using two different hydraulic equations; Supplementary Information) for streams and rivers is close to results of recent regional studies^{28,29} but is significantly higher than the value used in a recent global calculation¹⁰ and a calculation for the Amazon⁷. That value was not estimated systematically in the case of the former calculation and was not done before many measurements were available in the case of the latter. We also predict a decreasing gas transfer velocity with increasing stream order (Supplementary Table 1), which is consistent with recent field measurements²³. In a new metadata analysis of

whole-stream tracer releases in streams and small rivers, the average value was 4.7 m d^{-1} (ref. 29). These experiments, however, were limited to low discharge and, because turbulence is positively correlated with discharge, the values reported for small streams and rivers here are reasonable for average flow conditions. For large rivers, we predict a gas transfer velocity of $\sim 3\text{--}4 \text{ m d}^{-1}$ (Supplementary Table 1), which is also close to a recent synthesis for lowland rivers²⁸ that reported an average of 4.3 m d^{-1} and argued that many previous studies have probably underestimated k , which is generally higher in wet mountainous regions (Fig. 1).

We use two methods to estimate the gas transfer velocity for lakes and reservoirs. The first uses globally gridded wind speed and an empirical relationship between k_{600} , that is, the gas exchange velocity normalized to CO_2 at 20°C , and wind²¹ (Supplementary Information). The second uses new estimates of the gas transfer velocity for lakes of different sizes²², which assumes a primary role of fetch, that is, the distance travelled by the wind over the water, in regulating k in these systems. The wind speed and lake size models provided global average estimates of 0.74 and 1.33 m d^{-1} , respectively. Thus, a global average gas transfer velocity for lakes and reservoirs is approximately 1.0 m d^{-1} , which is much lower than the global average for streams and rivers (Fig. 2) but is consistent with a recent regional study³⁹.

Global CO_2 evasion from inland waters

Our estimated fluxes are lower than the most recent estimates for lakes and reservoirs but are higher for streams and rivers. For streams and rivers, we estimate a flux of 1.8 Pg C yr^{-1} . This is greater than in previous studies that have reported a stream and river evasion rate of $\sim 0.5\text{--}1 \text{ Pg C yr}^{-1}$ (refs 10–12), but is defensible considering stream and river evasion rates of 0.5 Pg C yr^{-1} from temperate regions⁹ and $\sim 0.6 \text{ Pg C yr}^{-1}$ from the Amazon^{7,8} alone. For lakes and reservoirs, our estimate of $\sim 0.3 \text{ Pg C yr}^{-1}$ is less than the most recent estimates of $\sim 0.5\text{--}0.6 \text{ Pg C yr}^{-1}$ (refs 10, 40) but is proximate to some of the older estimates^{12,41} (Supplementary Fig. 7). Our new estimate is less than more recent estimates owing to a smaller lake and reservoir area ($3 \times 10^6 \text{ km}^2$ compared with $4.2 \times 10^6 \text{ km}^2$); because we used the median instead of the mean as a representative value for the skewed distributions of p_{CO_2} , particularly in saline lakes; and because we account for generally lower p_{CO_2} in large lakes and reservoirs, which are important to the total area (Fig. 2).

There is a large amount of uncertainty associated with these estimates. We performed a Monte Carlo analysis to estimate the variance of our methodology by providing a distribution for the gas transfer velocity, the surface area and the dissolved CO_2 concentration for each COSCAT region, and then randomly sampled within these distributions for 1,000 iterations (Supplementary Information). The simulation predicted a flux of 1.8 Pg C yr^{-1} ($1.5\text{--}2.1 \text{ Pg C yr}^{-1}$; 5th and 95th confidence interval percentiles) for streams and rivers and $0.32 \text{ Pg C yr}^{-1}$ ($0.06\text{--}0.84 \text{ Pg C yr}^{-1}$; 5th and 95th confidence interval percentiles) for lakes and reservoirs. For streams and rivers, the uncertainty within COSCAT regions was positively correlated with the mean value of the flux, with regions with a high flux normalized to land area having the highest standard deviation (Supplementary Fig. 5). For lakes and reservoirs, the large range in the confidence interval is due to the nonlinear relationship between lake abundance and area and to uncertainty in the number or area of small lakes, which cannot at present be determined on the regional scale. In addition to the uncertainty estimated by the Monte Carlo analysis, there is considerable uncertainty in inland water science that may affect these estimates. Although we attempted to account for it in our analysis by using medians and adjusting the high range for the stream–river Monte Carlo analysis (Supplementary Information), there is still the potential that this method overestimates stream and river CO_2 as a result of potential biases and errors in calculating CO_2 from pH and alkalinity and as a result of the presence of organic acids (Supplementary Information). The overestimation of CO_2 is potentially affecting areas with few calculated CO_2 values and

high fluxes, such as Southeast Asia (Supplementary Information). Representative direct p_{CO_2} measurements are needed globally. In addition to improved CO_2 estimates, future research is needed on the distribution of lakes to refine estimates of lake area. Another large research gap is a lack of measurements of the gas transfer velocity of streams during average-to-high flows and in watersheds with a high slope. High-resolution global maps of stream length are still missing for the high latitudes. Further research on hydraulic relationships is needed, particularly in the tropics and high latitudes. For lakes, representative winter CO_2 measurements are missing and are often several times higher than during other seasons⁴². A further discussion on data limitations is provided in Supplementary Information.

A flux of 1.8 Pg C yr^{-1} for streams and rivers is large considering their small surface area, reinforcing the concept that streams and rivers are hotspots for exchange. Approximately 70% of the stream CO_2 evasion originates from waters located on only ~20% of the Earth's surface. Regions supporting this evasion include Southeast Asia, Amazonia, Central America, Europe, regions of South America west of the Andes, Southeast Alaska, small portions of western Africa and the eastern edge of East Asia (Fig. 1). Missing from this list is most of the northern high latitudes. The COSCAT drainages that include the Yenisei, Lena, Kolyma and Yana rivers, for instance, make up ~6% of the Earth's surface area but are responsible for only ~2% of global evasion. It is important to note that the surface area of rivers and streams in northern high latitudes is mainly extrapolated from empirical relationships between climate and percentage water cover at low latitudes (Supplementary Information), and that northern regions may have unique scaling laws and biogeochemistry that are currently not adequately understood. Thus, the evasion of CO_2 from northern latitudes needs further research. Africa, which is undersampled for CO_2 , also is predicted to make a low contribution, supporting only ~6% of annual CO_2 evasion despite making up ~22% of the terrestrial surface area.

This study further stresses the disproportionately high contribution of lower-order streams. We report a decrease in stream surface area and gas transfer velocity with increasing stream order (Supplementary Table 1). It is worth noting that the lower-order systems are undersampled for CO_2 , that they are not consistently gauged for discharge and that their surface area is difficult to measure directly by remote sensing. In this study, we were not able to assign CO_2 by stream order, but previous studies argue for a higher concentration of CO_2 in small streams and rivers^{9,19}. Further study on the surface area and CO_2 of small stream is needed.

For lakes and reservoirs, regions of high fluxes were estimated from the high latitudes and tropical regions (Fig. 2). We also conclude that ~50% of the emissions are from the world's largest lakes, owing to their large surface area and gas transfer velocity (Supplementary Information). However, large lakes are currently inadequately surveyed for both concentration and k . We also conclude that tropical lakes contribute disproportionately (Fig. 2), constituting only 2.4% of the global lake area but accounting for 34% of the global lake CO_2 emission, owing to high p_{CO_2} and high gas exchange rates. This could be due to the higher frequency of flooding of tropical lakes, which enhances terrestrial transfers. Lake CO_2 emissions per land area were highest in the humid tropics, but were also high in lake-rich boreal and Arctic regions (Fig. 2). Saline lakes, in contrast, are less important than previously reported⁴³, contributing ~18% to the global lake CO_2 evasion rather than ~50%. Much of this evasion is due to the Caspian Sea, the largest freshwater body on Earth, for which there are some calculated estimates of CO_2 but no proper survey.

The importance of the entire drainage network to CO_2 evasion provides information on the origins of inland water CO_2 . The high evasion rate in low-order streams is consistent with a large terrestrial soil CO_2 supply⁸, which could also be important to lake effluxes^{39,44}. The evasion of this CO_2 is, however, rapid⁴⁵ and cannot explain all of the evasion from higher-order systems and lakes and reservoirs. Although additional terrestrial soil CO_2 can still be added to these systems through

groundwater, contributions from organic matter decomposition and inorganic and organic carbon subsidies from fringing wetlands⁴⁶ are still needed to sustain a global CO_2 evasion rate of 2.1 Pg C yr^{-1} . The role of wetlands could be particularly important in hotspots such as Amazonia and southeast Asia⁴⁷. Systematic campaigns are needed to further evaluate the relative importance of different sources.

Understanding the relative importance of these sources is crucial to the global carbon budget. The evasion of terrestrial soil CO_2 in inland waters is part of terrestrial respiration. Although a 2.1 Pg C yr^{-1} displacement of global terrestrial net primary production (NPP) to inland waters represents only ~4% of NPP, the difference between terrestrial heterotrophic respiration and fires and NPP is of the order of ~1.5 Pg C yr^{-1} (ref. 48). Terrestrial approaches that attempt to determine this difference do not have the same ability to account for inland water evasion of CO_2 . A recent study demonstrated that ~1.2–2.2% of terrestrial NPP is evaded from lakes in catchments of England⁴⁴; thus, ignoring inland water CO_2 evasion could cause significant errors in regional-scale CO_2 budgets calculated using methods that rely on ecosystem-level CO_2 flux measurements. The percentage of evasion supported by the decomposition of terrestrial organic matter, added to the amount of terrestrial organic matter exported by rivers to the coastal ocean, also determines the total flux of terrestrial organic matter from the landscape, a flux that is not currently well constrained globally. Finally, if only a small percentage of this flux has an anthropogenic component it is important to the attribution of anthropogenic carbon in the global carbon budget^{49,50}.

METHODS SUMMARY

For inland waters, we relied almost exclusively on calculated CO_2 , determined from pH, alkalinity and temperature using PHREEQC version 2. Water chemistry data was taken from the literature and various governmental data sets and incorporated into the GloRiCh database. Data were collected and digitized over a period of ten years. For this analysis, 6,708 sampling locations were identified for streams and rivers and 25,699 single observations were made for lakes and reservoirs (Supplementary Information).

The surface area of inland waters was estimated using various geospatial products and scaling. For streams and rivers, we used HYDROSHEDS²⁴ and NHDPLUS to estimate length and hydraulic equations from the literature and USGS, along with global gridded run-off data²⁷ to estimate width. This could only be done for regions at latitudes below 60° N ; for higher latitudes, we used statistical relationships from regions below 60° N . For lakes and reservoirs, we used the Global Lakes and Wetlands Database for lakes more than 3.16 km^2 in area and used size distribution relationships from the literature^{16,33} to extrapolate to smaller lakes and reservoirs.

For streams and rivers, we estimated the gas transfer velocity using a recently published equation³⁰ that estimates k on the basis of slope and stream flow velocity. Stream flow velocity was estimated using hydraulic equations from the literature and USGS along with global gridded run-off data²⁷. Slope was determined using stream lines from HYDROSHEDS and elevation data from multiple sources (Supplementary Information). For lakes and reservoirs, we used two approaches for estimating the gas transfer velocity. The first used the relationship between k and wind speed given in ref. 21, whereas the second used the recently published relationship between lake area and k (ref. 22).

We calculated fluxes and tested the uncertainty of this efflux calculation using a Monte Carlo simulation (Supplementary Information).

Received 7 January; accepted 7 October 2013.

1. Le Quéré, C. *et al.* Trends in the sources and sinks of carbon dioxide. *Nature Geosci.* **2**, 831–836 (2009).
2. Kempe, S. Sinks of the anthropogenically enhanced carbon cycle in surface fresh waters. *J. Geophys. Res.* **89**, 4657–4676 (1984).
3. Park, P. K., Hager, S. W. & Cissell, M. C. Carbon dioxide partial pressure in the Columbia river. *Science* **166**, 867–868 (1969).
4. Raymond, P. A., Caraco, N. F. & Cole, J. J. Carbon dioxide concentration and atmospheric flux in the Hudson River. *Estuaries* **20**, 381–390 (1997).
5. Frankignoulle, M. *et al.* Carbon dioxide emission from European estuaries. *Science* **282**, 434–436 (1998).
6. Kling, G. W., Kipphut, G. W. & Miller, M. C. Arctic lakes and streams as gas conduits to the atmosphere: implications for tundra carbon budgets. *Science* **251**, 298–301 (1991).
7. Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M. & Hess, L. L. Carbon dioxide evasion from central Amazonian wetlands as a significant source of atmospheric CO_2 in the tropics. *Nature* **416**, 617–620 (2002).

8. Johnson, M. S. *et al.* CO₂ efflux from Amazonian headwater streams represents a significant fate for deep soil respiration. *Geophys. Res. Lett.* **35**, L17401 (2008).
9. Butman, D. & Raymond, P. A. Significant efflux of carbon dioxide from streams and rivers in the United States. *Nature Geosci.* **4**, 839–842 (2011).
10. Aufdenkampe, A. K. *et al.* Riverine coupling of biogeochemical cycles between land, oceans and atmosphere. *Front. Ecol. Environ.* **9**, 53–60 (2011).
11. Battin, T. J. *et al.* The boundless carbon cycle. *Nature Geosci.* **2**, 598–600 (2009).
12. Cole, J. J. *et al.* Plumbing the global carbon cycle: integrating inland waters into the terrestrial carbon budget. *Ecosystems* **10**, 172–185 (2007).
13. Richey, J. E. in *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World* (eds Field, C. B. & Raupach, M. R.) 329–341 (Scope 62, Island, 2004).
14. Downing, J. A. & Duarte, C. M. in *Encyclopedia of Inland Waters* 469–478 (Academic, 2009).
15. Lehner, B. & Doll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**, 1–22 (2004).
16. McDonald, C. P., Rover, J. A., Stets, E. G. & Striegl, R. G. The regional abundance and size distribution of lakes and reservoirs in the United States and implications for estimates of global lake extent. *Limnol. Oceanogr.* **57**, 597–606 (2012).
17. Buffam, I. *et al.* Integrating aquatic and terrestrial components to construct a complete carbon budget for a north temperate lake district. *Glob. Change Biol.* **17**, 1193–1211 (2011).
18. Christensen, T. R. *et al.* A catchment-scale carbon and greenhouse gas budget of a subarctic landscape. *Phil. Trans. R. Soc. A* **365**, 1643–1656 (2007).
19. Humborg, C. *et al.* CO₂ supersaturation along the aquatic conduit in Swedish watersheds as constrained by terrestrial respiration, aquatic respiration and weathering. *Glob. Change Biol.* **16**, 1966–1978 (2009).
20. Jonsson, A. *et al.* Integrating aquatic carbon fluxes in a boreal catchment carbon budget. *J. Hydrol.* **334**, 141–150 (2007).
21. Cole, J. J. & Caraco, N. F. Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF₆. *Limnol. Oceanogr.* **43**, 647–656 (1998).
22. Read, J. S. *et al.* Lake-size dependency of wind shear and convection as controls on gas exchange. *Geophys. Res. Lett.* **39**, L09405 (2012).
23. Striegl, R. G., Dornblaser, M. M., McDonald, C. P., Rover, J. R. & Stets, E. G. Carbon dioxide and methane emissions from the Yukon River system. *Glob. Biogeochem. Cycles* **26**, GB0E05 (2012).
24. Lehner, B., Verdin, K. & Jarvis, J. New global hydrograph derived from spaceborne elevation data. *Eos* **89**, 93–94 (2008).
25. Dodds, P. S. & Rothman, D. H. Geometry of river networks. I. Scaling, fluctuations, and deviations. *Physical Rev. E* **63**, 016115 (2001).
26. Leopold, L. B. & Maddock, T. *Geological Survey Professional Paper 252* 57 (US Government Printing Office, 1953).
27. Fekete, B. M., Vorosmarty, C. J. & Grabs, W. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Glob. Biogeochem. Cycles* **16**, 15 (2002).
28. Alin, S. R. *et al.* Physical controls on carbon dioxide transfer velocity and flux in low-gradient river systems and implications for regional carbon budgets. *J. Geophys. Res.* **116**, G01009 (2011).
29. Raymond, P. A. *et al.* Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers. *Limnol. Oceanogr. Fluids Environ.* **2**, 41–53 (2012).
30. Lauerwald, R., Hartmann, J., Moosdorf, N., Kempe, S. & Raymond, P. A. What controls the spatial patterns of the riverine carbonate system? A case study for North America. *Chem. Geol.* **337–338**, 114–127 (2013).
31. Meybeck, M., Durr, H. H. & Vorosmarty, C. J. Global coastal segmentation and its river catchment contributors: a new look at land-ocean linkage. *Glob. Biogeochem. Cycles* **20**, GB1S90 (2006).
32. Prigent, C., Papa, F., Aires, F., Rossow, W. B. & Matthews, E. Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *J. Geophys. Res.* **112**, D12107 (2007).
33. Downing, J. A. *et al.* Global abundance and size distribution of streams and rivers. *Inland Wat.* **2**, 229–236 (2012).
34. Downing, J. A. *et al.* The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **51**, 2388–2397 (2006).
35. Meybeck, M. in *Physics and Chemistry of Lakes* (eds Lerman, D., Imboden, M. & Gat, J. R.) 1–35 (Springer, 1995).
36. Cole, J. J. & Caraco, N. F. Carbon in catchments: connecting terrestrial carbon losses with aquatic metabolism. *Mar. Freshw. Res.* **52**, 101–110 (2001).
37. Abril, G. *et al.* Carbon dioxide and methane emissions and the carbon budget of a 10-year old tropical reservoir (Petit Saut, French Guiana). *Glob. Biogeochem. Cycles* **19**, GB4007 (2005).
38. Barros, N. *et al.* Carbon emission from hydroelectric reservoirs linked to reservoir age and latitude. *Nature Geosci.* **4**, 593–596 (2011).
39. McDonald, C. P., Stets, E. G. & Striegl, R. G. B. D. Inorganic carbon loading as a primary driver of dissolved carbon dioxide concentrations in lakes and reservoirs of the contiguous United States. *Glob. Biogeochem. Cycles* **27**, 285–295 (2013).
40. Tranvik, L. J. *et al.* Lakes and reservoirs as regulators of carbon cycling and climate. *Limnol. Oceanogr.* **54**, 2298–2314 (2009).
41. Cole, J. J., Caraco, N. F., Kling, G. W. & Kratz, T. K. Carbon dioxide supersaturation in the surface waters of lakes. *Science* **265**, 1568–1570 (1994).
42. Kortelainen, P. *et al.* Sediment respiration and lake trophic state are important predictors of large CO₂ evasion from small boreal lakes. *Glob. Change Biol.* **12**, 1554–1567 (2006).
43. Duarte, C. M. *et al.* CO₂ emissions from saline lakes: a global estimate of a surprisingly large flux. *J. Geophys. Res.* **113**, G04041 (2008).
44. Maberly, S. C., Barker, P. A., Stott, A. W. & De Ville, M. M. Catchment productivity controls CO₂ emissions from lakes. *Nature Clim. Change* **3**, 391–394 (2013).
45. Davidson, E. A., Figueiredo, R. O., Markewitz, D. & Aufdenkampe, A. K. Dissolved CO₂ in small catchment streams of eastern Amazonia: a minor pathway of terrestrial carbon loss. *J. Geophys. Res.* **115**, G04005 (2010).
46. Lauerwald, R., Hartmann, J., Ludwig, W. & Moosdorf, N. Assessing the nonconservative fluvial fluxes of dissolved organic carbon in North America. *J. Geophys. Res.* **117**, G01027 (2012).
47. Abril, G. *et al.* Amazon River carbon dioxide outgassing fuelled by wetlands. *Nature* <http://dx.doi.org/10.1038/nature12797> (in the press).
48. Sabine, C. L. *et al.* in *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World* (eds Field, C. B. & Raupach, M. R.) 17–45 (Scope 62, Island, 2004).
49. Raymond, P. A., Oh, N. H., Turner, R. E. & Broussard, W. Anthropogenically enhanced fluxes of water and carbon from the Mississippi River. *Nature* **451**, 449–452 (2008).
50. Regnier, P. *et al.* Anthropogenic perturbation of the carbon fluxes from land to ocean. *Nature Geosci.* **6**, 597–607 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements P.A.R. and M.H. were partly funded by a NASA grant (NNX11AH68G) to P.A.R. P.A.R. also received support from a fellowship from L-IPSL labex program. S.S. was supported by Formas. R.L. and J.H. were funded by the EU project GeoCarbon (U4603EUU1104) and by DFG (EXC 177 and DFG HA 4472/6-1). This represents a contribution to the RECCAP process. R.S. and D.B. are part of the Inland Water Science Group of the USGS LandCarbon Project.

Author Contributions P.A.R. designed and performed this analysis and wrote most of the paper. S.S. performed the lake and reservoir CO₂ and *k* analyses, and C.M. modelled lake and reservoir area data and provided material for these calculations for Supplementary Information. P.K. provided *p*CO₂ data and helped with lake analyses. R.L. and J.H. produced the global CO₂ data set. M.H. provided the GIS technical input. D.B. assisted with the GIS technical input and overall analysis and helped produce the figures. R.S. provided input on the use of USGS data and contributed to the overall analysis. E.M. provided COSCAT data on global discharge and dissolved organic carbon. H.D. provided COSCAT information and input on GIS analysis. P.K., C.H. and M.M. provided data for the lake CO₂ global data set. P.C. provided assistance with the sensitivity analysis and writing the final paragraph. P.G. provided data necessary to determine average watershed area for COSCAT regions. All authors read and commented on drafts of this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.A.R. (peter.raymond@yale.edu).

Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition

Miler T. Lee^{1*}, Ashley R. Bonneau^{1*}, Carter M. Takacs¹, Ariel A. Bazzini¹, Kate R. DiVito¹, Elizabeth S. Fleming¹ & Antonio J. Giraldez^{1,2}

After fertilization, maternal factors direct development and trigger zygotic genome activation (ZGA) at the maternal-to-zygotic transition (MZT). In zebrafish, ZGA is required for gastrulation and clearance of maternal messenger RNAs, which is in part regulated by the conserved microRNA miR-430. However, the factors that activate the zygotic program in vertebrates are unknown. Here we show that Nanog, Pou5f1 (also called Oct4) and SoxB1 regulate zygotic gene activation in zebrafish. We identified several hundred genes directly activated by maternal factors, constituting the first wave of zygotic transcription. Ribosome profiling revealed that *nanog*, *sox19b* and *pou5f1* are the most highly translated transcription factors pre-MZT. Combined loss of these factors resulted in developmental arrest before gastrulation and a failure to activate >75% of zygotic genes, including *miR-430*. Our results demonstrate that maternal Nanog, Pou5f1 and SoxB1 are required to initiate the zygotic developmental program and induce clearance of the maternal program by activating *miR-430* expression.

In animals, maternal gene products drive early development in a transcriptionally silent embryo, and are responsible for ZGA. ZGA occurs during the MZT, when developmental control transfers to the embryonic nucleus. This universal transition represents a major reprogramming event that requires (1) chromatin remodelling to provide transcriptional competency; (2) specific activation of a new transcriptional program; and (3) clearance of the previous transcriptional program. In *Drosophila*, maternal Zelda is required for activating the first zygotic genes through binding of TAGteam cis elements^{1,2}. However, the maternal factors that mediate ZGA in vertebrates remain largely unknown^{3,4}. In zebrafish, ZGA coincides with the midblastula transition (MBT) ~3 h post-fertilization (h.p.f.), during which genome competency is established through widespread changes in chromatin^{5,6} and DNA methylation^{7,8}. Bivalent chromatin marks are associated with zygotic genes thought to be 'poised' for activation⁵. Yet, many loci with active marks seem to be transcriptionally inactive⁵, indicating that competent genes require induction by additional factors. ZGA is required for epiboly⁹ and the clearance of maternal mRNAs, a process regulated in part by the conserved microRNA (miRNA) miR-430 (refs 10–12). Although significant advances have taken place in understanding how vertebrate embryos acquire transcriptional competency and orchestrate the clearance of the maternal program, the factors that control activation of the specific genes during ZGA remain unknown. Here we combine loss-of-function analyses, high-throughput sequencing and ribosome footprinting to identify factors that activate the first wave of zygotic transcription to initiate nuclear control of embryonic development.

Identifying the first zygotic transcripts

To define factors that mediate transcriptional activation, we first sought to identify the earliest genes transcribed from the zygotic genome. Accurate characterization of the early transcriptome faces two main challenges: (1) zygotic transcription of a gene can be masked by a large maternal contribution; and (2) poly(A)⁺ selection of mRNAs can lead

to apparent increases in gene expression, reflecting delayed polyadenylation of maternal mRNAs rather than transcription. We reasoned that maternal mRNAs are spliced during oogenesis, so examining introns from total RNA would allow us to quantify *de novo* transcription independent of polyadenylation or maternal contribution. We performed Illumina total RNA sequencing on wild-type embryos after the onset of zygotic transcription (4 h.p.f., sphere, and 6 h.p.f., shield) (Fig. 1a) compared to embryos before the MZT (2 h.p.f., 64-cell stage) and α -amanitin-treated embryos (assayed at sphere and shield), which lack zygotic transcription. This analysis identified 608 genes with significant increases in exon or intron expression levels >5 RPKM (reads per kilobase per million reads) at sphere stage ($P < 0.1$, Benjamini–Hochberg multiple test correction) (Fig. 1b, c and Extended Data Fig. 1a–h). Intron signal identifies an additional 6,602 genes with low levels of transcription by 4 h.p.f., and 9,330 transcribed genes by 6 h.p.f., expanding the number of zygotically expressed genes previously identified^{13,14} (Extended Data Fig. 1i–o and Supplementary Data 1). Over 74% of these are genes with maternal contributions (maternal and zygotic genes), most of which are only identified by elevated intron signal (Fig. 1b and Extended Data Fig. 1g), reflecting the sensitivity of this method to detect *de novo* transcription.

Next, we examined which genes are directly triggered by the maternal program in the 'first wave' of transcription by 4 h.p.f. versus those activated by zygotic factors. We reasoned that blocking zygotic gene function while leaving maternal factors unaffected would uncouple the first from subsequent waves of zygotic transcription. To this end, we inhibited splicing of zygotic mRNAs using morpholinos complementary to U1 and U2 spliceosomal RNAs (U1U2 MO) (Fig. 1d and Extended Data Fig. 1a–d)¹⁵. U1U2 MO embryos arrest before epiboly (Fig. 1a), despite remaining transcriptionally active. Illumina sequencing revealed an enrichment in intron–exon boundary reads (Fig. 1e) and activation of a subset of zygotic transcripts to levels >5 RPKM (Methods); these genes constitute the first wave of zygotic transcription (Fig. 1f). To test that these first-wave genes are indeed independent

¹Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA. ²Yale Stem Cell Center, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

*These authors contributed equally to this work.

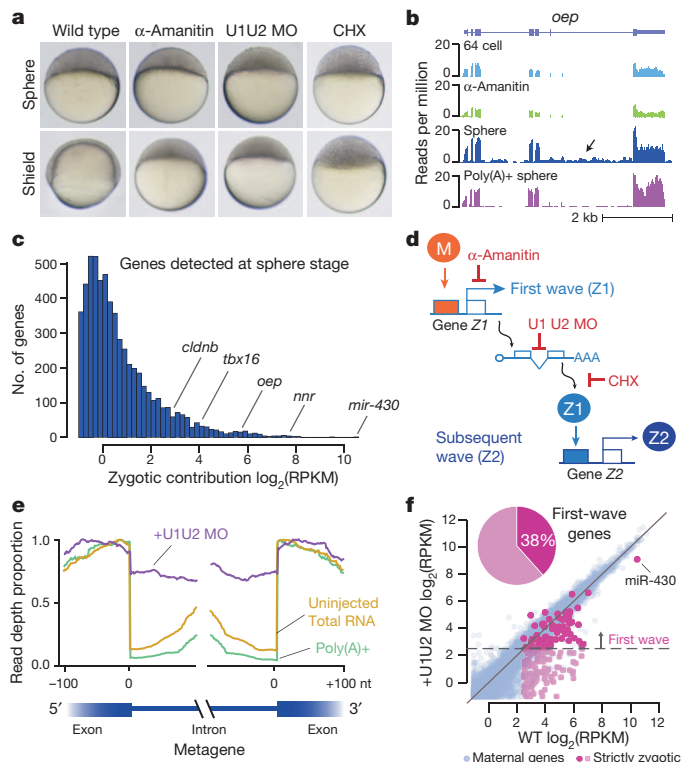


Figure 1 | Characterization of the zygotic transcriptome. **a**, Embryos showing the effects of α -amanitin, U1U2 morpholino (U1U2 MO) and cycloheximide (CHX). **b**, Sequencing read density across *oep*. Intronic signal increases with zygotic expression in total RNA. **c**, Expression histogram of zygotic genes. **d**, Maternal (M) but not zygotic factors (Z1) can activate transcription on splice or translation inhibition. **e**, Metagene of read density across exon–intron boundaries in first-wave genes. U1U2 MO shows enriched intron signal (purple). **f**, Biplot comparing expression in wild type and U1U2 MO. Points above 5 RPKM in U1U2 MO are considered first-wave genes.

of zygotic factors, we treated embryos with cycloheximide (CHX) before MBT (32-cell stage) to block translation of zygotic mRNAs selectively, while allowing translation of maternal mRNAs. CHX-treated embryos also fail to reach epiboly (Fig. 1a) and have a highly correlated transcriptome profile with U1U2 MO (Pearson's $R = 0.97$, Extended Data Fig. 2), confirming first-wave transcription in the absence of zygotic proteins. First-wave genes comprise both embryonic-specific and house-keeping genes ubiquitously expressed in adult tissues (Extended Data Fig. 3a) and are enriched in pattern specification, gastrulation and chromatin modifying functions (Extended Data Fig. 3b). We validated a subset of these genes by RT–PCR, including *kif4b*, *nnr* and *isg15* (Extended Data Fig. 3c–k). Notably, the pri-miR-430 polycistron is highly expressed as part of this first wave ($>1,000$ RPKM) (Fig. 1c, f). Together, these results identify 269 first-wave genes expressed by sphere stage for which maternal factors are sufficient for activation.

Nanog, SoxB1 and Pou5f1 activate the first wave

Considering the specific, widespread and steep pattern of zygotic gene activation, we proposed that the factors that trigger the first wave may include sequence-specific transcriptional regulators highly translated before ZGA. We analysed the translation levels of all maternal mRNAs using ribosome profiling data (Fig. 2a)¹⁶. We found that Nanog, Sox19b and Pou5f1 are the most highly translated sequence-specific transcription factors in the pre-MZT transcriptome (Fig. 2b). Pou5f1, the SoxB1 family (which includes Sox2 and Sox19b) and Nanog are key transcription factors involved in maintaining pluripotency in embryonic stem (ES) cells (reviewed in refs 17, 18). In zebrafish, Pou5f1 provides temporal control of gene expression¹⁹ and together with SoxB1 regulates dorsal–ventral patterning and neuronal development^{18,20–23}, whereas Nanog is essential for endoderm formation through regulation of zygotic *mtx2* (ref. 24).

To examine the roles of Nanog, Sox19b and Pou5f1 in activating zygotic gene expression, we combined a maternal–zygotic loss-of-function (LOF) Pou5f1 (*MZpou5f1*)²¹ with previously published translation-blocking morpholinos for Nanog (ref. 24) and SoxB1 (ref. 20) (Methods). Because Sox2, Sox3 and Sox19a have been shown to compensate for

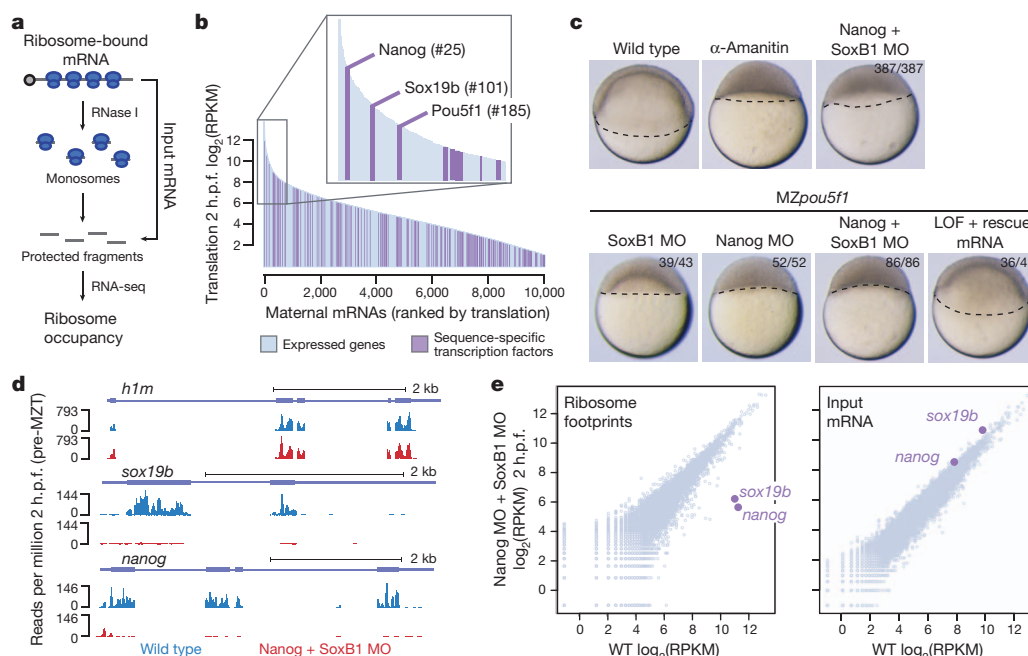


Figure 2 | Identification of Nanog, SoxB1 and Pou5f1 as zygotic gene regulators. **a**, Schematic diagram illustrating ribosome profiling. **b**, Rank plot showing translation levels pre-MZT. Sequence-specific transcription factors are highlighted. **c**, Embryos with combined loss of Nanog plus SoxB1, Nanog plus Pou5f1 or triple LOF arrest similar to α -amanitin and are rescued with

nanog, *soxB1* and *pou5f1* mRNA injection. **d**, Ribosome footprints for *h1m*, *sox19b* and *nanog* in wild type and Nanog MO plus SoxB1 MO. *sox19b* and *nanog* are highly depleted in the morpholino condition. **e**, Biplots comparing wild-type and morpholino ribosome footprints and input mRNA.

Sox19b loss, we used a combination of morpholinos targeting all four *sox* genes²⁰ (Extended Data Fig. 4a). Simultaneous Nanog LOF in combination with SoxB1 or Pou5f1 resulted in complete developmental arrest before gastrulation, with >95% of the treated embryos failing to initiate epiboly ($n = 387$ and $n = 52$, respectively) (Fig. 2c and Extended Data Fig. 4b–e). This phenotype resembles that of α -amanitin-injected embryos, indicating that these factors have a role in activating zygotic genes. We used two different approaches to analyse the activity and specificity of these morpholinos. First, we performed ribosome profiling on wild-type and Nanog plus SoxB1 morpholino-injected embryos pre-MBT^{16,25}. Translation efficiency for both Nanog and Sox19b was reduced >97% in the morpholino-injected embryos compared to wild type (Fig. 2d and Extended Data Fig. 4f), but was largely unaffected for the rest of the transcriptome (Fig. 2e). Second, we co-injected mRNAs encoding Nanog and SoxB1 with the morpholinos and were able to rescue gastrulation (Fig. 2c and Extended Data Fig. 4c–e). Together, these results show that Nanog, Sox19b and Pou5f1 regulate progression through zygotic development and gastrulation.

Illumina sequencing revealed that combined loss of Nanog, SoxB1 and Pou5f1 results in widespread reduction in first-wave gene expression by 4 h.p.f.: 77% for strictly zygotic genes, 50% for maternal and zygotic genes (Fig. 3a, b and Extended Data Fig. 5). By 6 h.p.f., expression loss is systemic, with 86% of strictly zygotic and 79% of maternal and zygotic genes failing to be expressed to wild-type levels (Fig. 3a, b and Extended Data Fig. 5), an effect that was rescued by injection of the cognate mRNAs (Fig. 3c and Extended Data Figs 5 and 6). Comparing the single and double LOF transcriptomes to the triple, we found that regulation is often combinatorial and redundant, with Nanog LOF having the strongest effect and SoxB1 the weakest (Fig. 3d and Extended Data Fig. 7a–c). By 6 h.p.f., affected genes include housekeeping genes, general transcription factors (for example, *gata6*, *otx1*, *irx1b*, *ntla*) and major signalling components in gastrulation, anterior–posterior axis and dorsal–ventral axis specification (for example, *oep*, *fgf3*, *wnt11*, *chd*, *nog1*, *ndr2*, *bmp2b*) (Extended Data Fig. 7d, e). Together, these results show that Nanog, Pou5f1 and SoxB1 have a fundamental role in activating the first wave, an effect that propagates to subsequent waves resulting in a global impact on zygotic gene expression.

miR-430 is strongly activated by Nanog

Notably, among the first-wave genes co-regulated by Nanog, Pou5f1 and SoxB1 was miR-430, a miRNA family that functions in the clearance of maternal mRNAs in zebrafish and *Xenopus*^{10–12}. Northern blot analysis revealed a strong reduction of mature miR-430 levels in Nanog LOF embryos (Fig. 4a). Although individual loss of SoxB1 or Pou5f1 had no detectable effect on miR-430 expression, when combined with Nanog LOF they reduced miR-430 levels even further, a phenotype that was rescued by co-injecting the respective mRNAs (Fig. 4a–c). Nanog morpholino embryos failed to repress a GFP reporter of endogenous miR-430 activity²⁶, consistent with Nanog's role in activating miR-430 (Extended Data Fig. 8a, b).

To determine whether Nanog specifically binds the *miR-430* genomic locus, we analysed Nanog chromatin immunoprecipitation sequencing (ChIP-seq) data at high (3.3 h.p.f.) and dome stage (4.3 h.p.f.)²⁴. Consistent with widespread Nanog regulation, 74% of first-wave genes are bound by Nanog, a significant enrichment compared to subsequent-wave genes (Fig. 4d and Extended Data Fig. 9a). miR-430 is expressed from a 17-kilobase (kb) genomic region on chromosome 4, which includes 55 repeated miR-430 hairpin sequences. Because this locus is repetitive, it had been excluded from previous analyses; however, the sequences are largely unique relative to the rest of the genome. Reads aligning the *miR-430* locus were enriched >16-fold in the Nanog immunoprecipitation compared to whole-cell extract (Fig. 4e), indicating that strong Nanog binding throughout the locus correlates with strong miR-430 expression at ZGA. When the reads were aligned to the presumptive 5' end of the polycistron, we observed a strong peak of binding in a

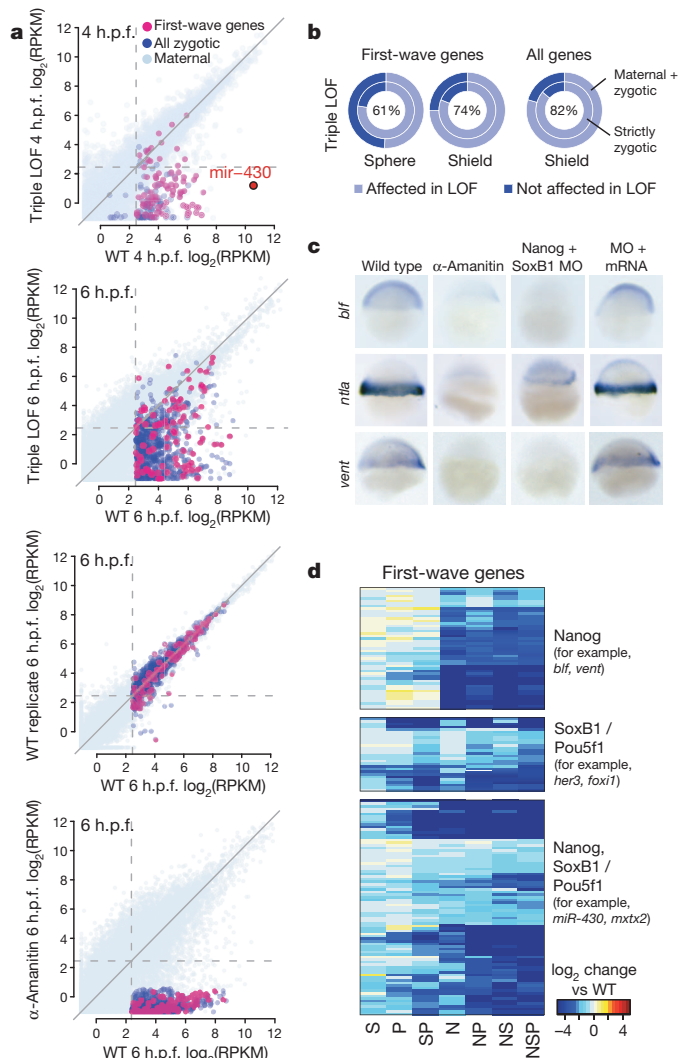


Figure 3 | Transcriptome-wide effects of loss of Nanog, SoxB1 and Pou5f1. **a**, Biplots showing widespread gene expression loss in the triple LOF at 4 and 6 h.p.f. **b**, Plots showing global effects of LOF. Percentages show the combined effect for strictly zygotic and maternal plus zygotic gene groups. **c**, *In situ* hybridization shows expression defects in LOF embryos, which are rescued by mRNA injection. **d**, Heat-map showing first-wave zygotic genes in single and combined LOF conditions. N, Nanog MO; P, MZ*pou5f1*; S, SoxB1 MO. Patterns shown are regulation by Nanog predominantly (top); SoxB1 and Pou5f1 (middle); or Nanog in combination with SoxB1 and Pou5f1 (bottom).

~600-nucleotide region between two miR-430 precursors, which contains three canonical Nanog binding sites (CATT[T/G][T/G]CA)^{24,27}.

To determine whether Nanog induces clearance of maternal mRNAs through activation of miR-430, we analysed the expression of an endogenous miR-430 target, *cd82b* (ref. 10). *cd82b* mRNA is maternally deposited and cleared in wild type by 6 h.p.f. (Fig. 5a). In contrast, *cd82b* mRNA is stabilized in MZ*dicer* mutants or α -amanitin-treated embryos, which lack miR-430 processing and expression, respectively. Similar loss of regulation is observed in Nanog plus SoxB1 MO, as well as triple LOF embryos, a defect that is rescued by providing the cognate mRNAs (Fig. 5b and Extended Data Fig. 8c). To determine the global effect of this regulation, we examined RNA-seq levels of maternal mRNAs containing miR-430 target sites. Loss of Nanog alone or in combination with loss of SoxB1 and MZ*pou5f1* resulted in miR-430 target stabilization, similar to MZ*dicer*^{10,16,26} (Fig. 5c and Extended Data Fig. 8d–f) ($P < 1 \times 10^{-51}$, two-sided Wilcoxon rank-sum test). A significant, but weaker, effect was observed in Pou5f1 plus SoxB1 LOF embryos

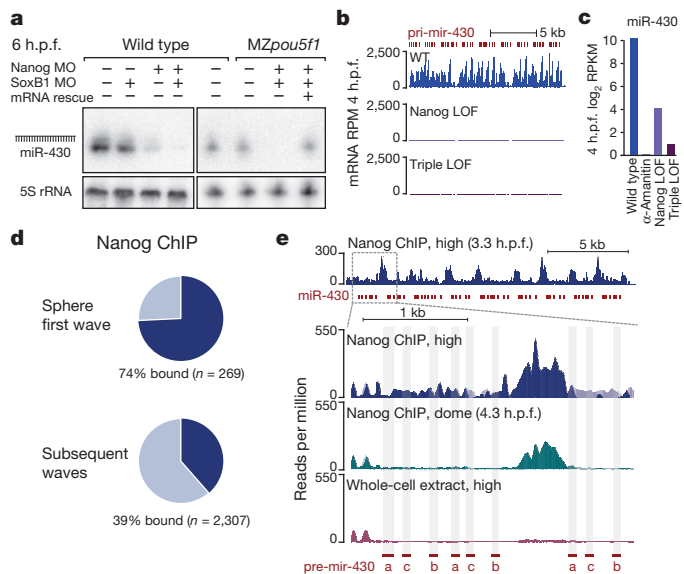


Figure 4 | miR-430 expression is regulated by Nanog. **a**, Northern blot shows that miR-430 is severely reduced in Nanog LOF and nearly undetectable in the triple LOF embryos. **b**, RNA-seq read levels of the pri-miR-430 polycistron in wild-type and LOF embryos. **c**, Bar plot of total miR-430 aligning reads. **d**, First-wave genes are highly bound by Nanog. **e**, Nanog binding across the miR-430 region (top panel) and a zoomed region where reads are preferentially aligned to the 5' end (bottom). Binding profiles show a strong peak between two precursors. pre-miR-430a, pre-miR-430b and pre-miR-430c are marked in red.

($P < 1 \times 10^{-25}$) (Extended Data Fig. 8d). These results show that Nanog together with Pou5f1 and SoxB1 activate miR-430 expression, thus revealing a genetic network that links maternal regulation of zygotic gene expression to zygotic clearance of maternal mRNAs.

Discussion

Our transcriptome analysis during the maternal-to-zygotic transition provides three major insights. First, maternal factors directly regulate hundreds of mRNAs that constitute the first wave of zygotic transcription. These targets are activated in the absence of zygotic gene function and are enriched for genes that guide early embryonic development. Transcriptional competence coincides with changes in the chromatin and DNA methylation states of the genome^{4–8}. Modifications to the

epigenetic landscape during the MZT may be sufficient to allow basal levels of transcription; however, we show here that maternal transcription factors have a vital role in shaping transcriptional output.

Second, we observe that Nanog, SoxB1 and Pou5f1, previously implicated in the maintenance of pluripotency, contribute to widespread activation of zygotic genes during the MZT. These maternal factors enhance transcriptional activation of more than 74% of first-wave zygotic genes, and by 6 h.p.f. influence expression of >80% genes overall. Simultaneous removal of Nanog with SoxB1 and/or Pou5f1 results in complete block of gastrulation and developmental arrest, similar to global inhibition of zygotic gene expression (Fig. 2c and Extended Data Fig. 9c). Nanog binds 74% of first-wave genes during the early stages of ZGA (Fig. 4d). Additionally, while this manuscript was under review, Pou5f1 and Sox2 were also shown to associate with ~40% of early zygotic genes²⁸. However, SoxB1 plus Pou5f1 LOF is insufficient to block gastrulation and zygotic development²⁸ (Fig. 2c). This highlights the central role of Nanog, which together with Pou5f1 and SoxB1 initiates the zygotic program of development, although it is likely that additional factors cooperate with them to provide genome competency and regulate the timing of ZGA⁴. These factors' role in vertebrates may be comparable to Zelda in *Drosophila*, in activating a large cohort of zygotic genes². In mouse, Oct4 and Nanog have been proposed to regulate gene expression at the 2-cell stage^{29,30} and along with Sox2 are required for specification of the blastocyst lineages^{31–33}. In fact, when we analyse early zygotic genes in mouse, we find that they are enriched for Nanog, Oct4 and Sox2 binding in embryonic stem cells (Extended Data Fig. 9b). Conceptually and mechanistically, many parallels exist between the MZT and the cellular reprogramming that occurs in induced pluripotent stem cells (iPS cells)^{3,12}. Indeed, reprogramming of terminally differentiated cells was first shown in the context of the early embryo through nuclear transfer^{34,35}. The onset of zygotic development can be viewed as a major reprogramming event that occurs on fusion of two terminally differentiated cells (sperm and oocyte). As shown in ES cells and iPS cells, Pou5f1, Nanog and Sox2 are central players in the induction^{36–40} and maintenance^{41–43} of pluripotency *in vivo* and *in vitro*^{17,35}. In these contexts, part of their role is to serve as 'pioneering' factors, binding to silent chromatin to facilitate *de novo* gene expression⁴⁴. We propose that this pioneering activity is recapitulated during the MZT, where an endogenous function of Nanog, SoxB1 and Pou5f1 is to mediate activation of the first wave of zygotic genes, establishing a transient pluripotent state.

Third, we show that Nanog together with SoxB1 and Pou5f1 directly regulate miR-430, which is responsible for clearance of maternal mRNAs during the MZT^{10–12}, facilitating the transfer of developmental control to the zygotic program (Extended Data Fig. 9c). Members of the conserved miR-430/295/302/372 family of miRNAs stabilize self-renewal fate in ES cells and enhance reprogramming efficiency^{45,46}. We propose that in both cases these miRNAs are 'clearing the slate' by accelerating the removal of mRNAs from the previous program, thus facilitating the establishment of new states by reprogramming factors¹². The marked upregulation of miR-430 expression by Nanog, SoxB1 and Pou5f1 provides a central link between the mechanisms that drive zygotic gene activation and the clearance of the previous maternal history.

METHODS SUMMARY

MZpou5f1^{hi349Tg/hi349Tg} and *MZdicer*^{hu896/hu896} were generated as previously described^{21,26}. All injections were performed at the one-cell stage. For translation inhibition, 32-cell stage embryos were incubated in media with 50 $\mu\text{g ml}^{-1}$ cycloheximide (Sigma Aldrich) at 28 °C until collection. Total RNA libraries were constructed using the TruSeq Stranded and Ribo-Zero Gold kits (Epicentre). Aligned reads were intersected with Ensembl r70 and RefSeq gene exon and intron annotations. Differential expression was performed using DESeq¹⁷. ChIP-seq data were analysed as described previously²⁴, except for the *miR-430* locus, for which unique alignments were not required. Ribosome profiling was performed as described in ref. 16, using the Epicentre ARTseq kit. Sequencing samples are summarized in Extended Data Table 1.

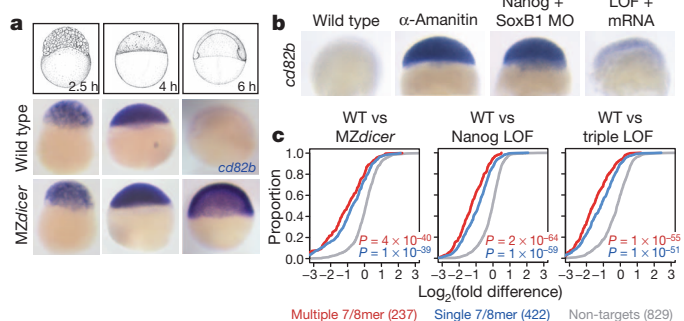


Figure 5 | miR-430 activity is abrogated by Nanog LOF. **a**, *In situ* hybridization showing degradation of miR-430 target *cd82b* at 6 h.p.f. in wild type, compared to stabilization in *MZdicer* (lacking miR-430 activity). **b**, *cd82b* is stabilized in the Nanog-SoxB1 LOF embryo, indicating loss of miR-430 activity. The effect is rescued by injection of *nanog* and *soxB1* mRNA. **c**, Cumulative plots showing stabilized expression of miR-430 targets in *MZdicer* and LOF embryos, compared to wild type. *P* values are for two-sided Wilcoxon rank-sum tests comparing each miR-430 target group to non-targets.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 May; accepted 4 September 2013.

Published online 22 September 2013.

- ten Bosch, J. R., Benavides, J. A. & Cline, T. W. The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. *Development* **133**, 1967–1977 (2006).
- Liang, H. L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456**, 400–403 (2008).
- Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136**, 3033–3042 (2009).
- Andersen, I. S. *et al.* Epigenetic marking of the zebrafish developmental program. *Curr. Top. Dev. Biol.* **104**, 85–112 (2013).
- Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464**, 922–926 (2010).
- Lindeman, L. C. *et al.* Prepatterned of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21**, 993–1004 (2011).
- Potok, M. E., Nix, D. A., Parnell, T. J. & Cairns, B. R. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153**, 759–772 (2013).
- Jiang, L. *et al.* Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* **153**, 773–784 (2013).
- Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119**, 447–456 (1993).
- Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**, 75–79 (2006).
- Lund, E., Liu, M., Hartley, R. S., Sheets, M. D. & Dahlberg, J. E. Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA* **15**, 2351–2363 (2009).
- Giraldez, A. J. microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. *Curr. Opin. Genet. Dev.* **20**, 369–375 (2010).
- Harvey, S. A. *et al.* Identification of the zebrafish maternal and paternal transcriptomes. *Development* **140**, 2703–2710 (2013).
- Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **21**, 1328–1338 (2011).
- Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).
- Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–237 (2012).
- Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136**, 2311–2322 (2009).
- Onichtchouk, D. Pou5f1/oct4 in pluripotency control: insights from zebrafish. *Genesis* **50**, 75–85 (2012).
- Onichtchouk, D. *et al.* Zebrafish Pou5f1-dependent transcriptional networks in temporal control of early development. *Mol. Syst. Biol.* **6**, 354 (2010).
- Okuda, Y., Ogura, E., Kondoh, H. & Kamachi, Y. B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. *PLoS Genet.* **6**, e1000936 (2010).
- Lunde, K., Belting, H. G. & Driever, W. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. *Curr. Biol.* **14**, 48–55 (2004).
- Hauptmann, G. *et al.* spiel ohne grenzen/pou2 is required for zebrafish hindbrain segmentation. *Development* **129**, 1645–1655 (2002).
- Belting, H. G. *et al.* Pou5f1 contributes to dorsoventral patterning by positive regulation of *vox* and modulation of *fgf8a* expression. *Dev. Biol.* **356**, 323–336 (2011).
- Xu, C. *et al.* Nanog-like regulates endoderm formation through the Mxt2-Nodal pathway. *Dev. Cell* **22**, 625–638 (2012).
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
- Giraldez, A. J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**, 833–838 (2005).
- Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
- Leichsenring, M., Maes, J., Mossner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**, 1005–1009 (2013).
- Foygel, K. *et al.* A novel and critical role for Oct4 as a regulator of the maternal-embryonic transition. *PLoS ONE* **3**, e4109 (2008).
- Tan, M. H. *et al.* An Oct4-Sall4-Nanog network controls developmental progression in the pre-implantation mouse embryo. *Mol. Syst. Biol.* **9**, 632 (2013).
- Keramari, M. *et al.* Sox2 is essential for formation of trophectoderm in the preimplantation embryo. *PLoS ONE* **5**, e13952 (2010).
- Frum, T. *et al.* Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Dev. Cell* **25**, 610–622 (2013).
- Messerschmidt, D. M. & Kemler, R. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev. Biol.* **344**, 129–137 (2010).
- Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182**, 64–65 (1958).
- Gurdon, J. B. & Melton, D. A. Nuclear reprogramming in cells. *Science* **322**, 1811–1815 (2008).
- Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
- Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
- Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biol.* **9**, 625–635 (2007).
- Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372–376 (2000).
- Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
- Judson, R. L., Babiarz, J. E., Venere, M. & Bloloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature Biotechnol.* **27**, 459–461 (2009).
- Subramanyam, D. *et al.* Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature Biotechnol.* **29**, 443–448 (2011).
- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Kamachi, W. Driever, A. F. Schier, N. Ivanova and I.-H. Park for reagents and fish lines; S. Mane and J. Overton for outstanding sequencing support; E. Zdobnov for hosting genomics data; A. Hubaud, N. Darricarrere and M. Koziol for contribution in the early stages of this project; and all the members of the Giraldez laboratory for discussions. This work was supported by NIH grants F32HD071697-02 (M.T.L.), T32GM007499 (A.R.B.), F32HD061194-03 (C.M.T.), Pew Fellows Program in Biomedical Sciences (A.A.B.), R01GM081602-06, R01GM103789-01, R01HD074078-02, the Pew Scholars Program in the Biomedical Sciences and the Yale Scholars Program (A.J.G.).

Author Contributions M.T.L., A.R.B. and A.J.G. designed the project, performed experiments and data analysis. M.T.L., A.R.B., C.M.T. and A.J.G. wrote the manuscript. C.M.T. designed and performed the cycloheximide experiment and contributed to *in situ* hybridizations. A.A.B. designed and performed ribosome profiling and U1U2 experiments. K.R.D. and E.S.F. assisted with gene validation.

Author Information Sequencing data are deposited in the Gene Expression Omnibus (GEO) database with accession number GSE47558. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.J.G. (antonio.giraldez@yale.edu).

METHODS

Zebrafish maintenance. *MZpou5f1*^{hi349Tg/hi349Tg} (ref. 48) were generated as previously described²¹. Embryos obtained from natural crosses between homozygous *MZpou5f1*^{hi349Tg/hi349Tg} mutants were injected with 30 pg of *pou5f1* mRNA at the one-cell stage. *MZdicer*^{hu896/hu896} fish were generated as described previously²⁶. Zebrafish wild-type embryos were obtained from natural crosses of TU-AB and TLF strains of mixed ages (5–17 months). Selection of mating pairs was random from a pool of 60 males and 60 females allocated for a given day of the month. Fish lines were maintained in accordance with AAALAC research guidelines, under a protocol approved by Yale University IACUC.

Treatments and mRNA injection. Embryos from all wild-type crosses were pooled following collection and distributed equally between experimental conditions. Unless otherwise stated, a minimum of 30 wild-type embryos were subjected to each treatment in each experimental replicate. Morpholinos were obtained from Gene Tools and re-suspended in nuclease-free water. Unless otherwise stated, 1 nl of morpholino solution was injected into dechorionated embryos at the one-cell stage. A combination of two morpholinos was used to target each gene in a 1:1 ratio as described in ref. 20, with one SoxB1 morpholino targeting a conserved region of both *sox2* and *sox3*. Nanog and SoxB1 morpholinos were previously described in refs 20, 24, respectively. For individual and combinatorial loss of function, wild-type and *MZpou5f1* embryos were injected with 1 ng of each SoxB1 morpholino (0.125 mM each) and 5 ng of Nanog morpholino (0.6 mM each). For inhibition of splicing, one morpholino (1.25 mM each) complementary to U1 and two morpholinos (0.6 mM each) complementary to isoforms of U2 spliceosomal RNAs (U1U2) were used^{15,49,50}. Divergence of the U2 genes in zebrafish requires the use of two different morpholinos to block activity.

Zebrafish Nanog and SoxB1 capped mRNA was generated by *in vitro* transcription using mMessage mMachine Sp6 kit (Ambion) in accordance with the manufacturer's instructions. For Nanog morpholino rescue, zebrafish *nanog* was cloned into a pCS2 vector and sense mutations introduced during PCR amplification (indicated in lowercase): 5'-ATGGCaGATGGAAaATGCCgGTGAGTTAC-3'. SoxB1 rescue constructs were provided by Y. Kamachi²⁰. To rescue the loss-of-function phenotype, 50 pg of Nanog and 20 pg of SoxB1 (5 pg each) mRNAs were injected either individually or together into morpholino-injected embryos at the one-cell stage. Triple loss-of-function embryos were additionally injected with 30 pg of *pou5f1* mRNA⁵¹.

For polymerase II inhibition, α -amanitin was obtained from Sigma Aldrich and re-suspended in nuclease-free water. Dechorionated embryos were injected with 0.2 ng of α -amanitin at the one-cell stage³².

For translation inhibition, wild-type embryos were collected and dechorionated at the one-cell stage. To allow for translation of maternal mRNAs, at 32-cell stage, embryos were transferred to media containing cycloheximide (50 μ g ml⁻¹) (Sigma Aldrich) and incubated at 28 °C. Embryos were collected and frozen in liquid nitrogen at sphere and shield stage. Total RNA was extracted from ten embryos using Trizol (Invitrogen) and re-suspended in 10 μ l RNase-free water.

To assay miR-430 activity, a GFP reporter was used as previously described²⁶. GFP and dsRed mRNAs were *in vitro* transcribed using mMessage mMachine Sp6 kit (Ambion) in accordance to the manufacturer's instructions. Embryos were injected with 150 pg of GFP reporter and 100 pg of dsRed loading control at the one-cell stage.

All phenotypes were initially assayed by one experimenter and blindly confirmed and/or imaged by another. Distribution-free statistics were used to determine significance, except for calculating RNA-seq differential expression (see below).

In situ hybridization. Template for *in situ* probes was amplified from shield stage cDNA and a T7-promoter sequence added for *in vitro* transcription. Primers are listed below. Antisense digoxigenin (DIG) RNA probes were generated by *in vitro* transcription in 20 μ l reactions consisting of 100 ng purified PCR product (8 μ l), 2 μ l DIG RNA labelling mix (Roche), 2 μ l \times 10 transcription buffer (Roche), and 2 μ l T7 RNA polymerase (Roche) in RNase-free water and purified using a Qiagen RNEasy kit. *In situ* protocol was followed as detailed previously²⁶. To reduce variability, the following conditions were combined in the same tube during *in situ* hybridization and recognized based on their morphology: (1) wild-type and α -amanitin-injected embryos and (2) Nanog plus SoxB1 MO with and without rescue mRNA. Before photo documentation, embryos were cleared using a 2:1 benzyl benzoate:benzyl alcohol solution. Images were obtained using a Zeiss stereo Discovery V12.

Northern analysis. To detect endogenous miR-430, ten wild-type and *MZpou5f1* embryos injected with Nanog morpholino and SoxB1 morpholino mix were collected at 6 h.p.f. and flash frozen in liquid nitrogen. Total RNA was extracted using Trizol (Invitrogen) and re-suspended in 5 μ l RNase-free water and 5 μ l \times 2 loading buffer (8 M urea, 50 mM EDTA, 0.2 mg ml⁻¹ xylene cyanol, and 0.2 mg ml⁻¹ bromophenol blue). Northern protocol was followed as detailed previously¹⁶.

Ribosome profiling. Fifty wild-type embryos injected with 1 nl of Nanog morpholino (0.6 mM each) and SoxB1 morpholino (0.125 mM each) mix and fifty non-injected embryos were collected at the 64-cell stage. Embryos were lysed using 800 μ l of a mammalian cell lysis buffer containing 100 μ g ml⁻¹ cycloheximide as per the manufacturer's instruction (ARTseq Ribosome Profiling kit, RPHMR12126, Epicentre). For nuclease treatment, 3 μ l of ARTseq nuclease was used. Ribosome protected fragments were run and 28–29-nt fragments were gel purified as previously described¹⁶ and cloned according to the manufacturer's protocol (ARTseq kit). Illumina libraries were constructed and sequence reads analysed as in ref. 16. Subsequent to sequencing, traces of exogenous RNA corresponding to a *nanog* antisense probe, and *ntla* sense and antisense, were detected outside the expected size range. Only 28- and 29-nt sense sequences were used in the analysis matching the size of the ribosome footprint.

Reverse transcription PCR (RT-PCR). Total RNA from ten embryos was extracted using Trizol (Invitrogen) at sphere and shield stage for each experimental condition. RNA was treated with TURBO DNase (Ambion) for 30 min at 37 °C and extracted using phenol chloroform. cDNA was generated by reverse transcription with random hexamers using SuperscriptII (Invitrogen). RT-PCR reactions were carried out at an annealing temperature of 60 °C for 35 cycles. Primers are listed below.

Illumina sequencing. Total RNA was extracted as above, and strand-specific TruSeq Illumina RNA sequencing libraries were constructed by the Yale Center for Genome Analysis. Before sequencing, samples were treated with Epicentre Ribo-Zero Gold kits according to the published protocol, to deplete ribosomal RNA. Samples were multiplexed on Illumina HiSeq 2000/2500 machines to produce single-end 76-nt reads. Sequencing samples are summarized in Extended Data Table 1.

Raw reads were initially filtered by aligning permissively to a ribosomal DNA index using Bowtie v0.12.9⁵³ with switches -seedlen 25 -n 3 -k 1 -y -e 10000. Unaligned reads were then aligned to the zebrafish Zv9 (UCSC danRer7) genome sequence using Tophat v2.0.7⁵⁴ with default parameters.

Hybrid gene models were constructed from the union of zebrafish Ensembl r70, RefSeq annotations (downloaded from <http://www.genome.ucsc.edu> on 8 February 2013) and Ensembl RNA-seq gene models⁵⁵. All overlapping transcript isoforms were merged to produce maximal exonic annotations. To quantify exonic expression levels per gene, genome-uniquely aligning reads overlapping ≥ 10 nt to the exonic region of a given gene were summed. To quantify intronic expression levels per gene, an annotation mask was first created consisting of repetitive sequences as annotated by RepeatMasker in addition to any region aligned by ≥ 2 reads in the α -amanitin samples; this is to minimize false-positive introns due to annotation inconsistencies, under the assumption that the transcriptionally inhibited α -amanitin transcriptome should contain no intron-containing transcripts. Valid intron-overlapping reads aligned the intronic region uniquely and overlapped no more than 50% to the masked regions. For the purposes of RPKM normalization, we considered intron length to be the number of unmasked nucleotides. We additionally identified reads that mapped to at most two different genic loci (for example, two closely related paralogues) and from these calculated 'meta gene' expression values. Meta genes were treated as conventional genes for differential expression, but counted as two different genes in subsequent analyses.

The *miR-430* locus is internally repetitive; therefore, reads were aligned to *miR-430* in a separate step using Bowtie with switches -n 2 -k 1 on the genomic region chr4:27999472-28021845, which spans the presumed *miR-430* polycistron. Reads overlapping any of the Ensembl annotated *miR-430* hairpins in this region were counted as *miR-430* cluster reads. Reads are counted only once, regardless of the number of times they overlap.

Differential gene expression analysis. Differential expression analysis was performed using the R package DESeq⁴⁷ with the parameters fit-type = local and sharingMode = fit-only. For exonic expression comparisons, raw exon-overlapping read counts were assembled for all genes with a raw read count of at least 10 in one or more of the samples. Genes annotated as Ensembl biotypes 'IG_C_pseudogene', 'IG_pseudogene', 'IG_V_pseudogene', 'misc_RNA', 'Mt_rRNA', 'Mt_tRNA', 'non_coding', 'nonsense_mediated_decay', 'retained_intron', 'rRNA', 'sense_intronic', 'sense_overlapping', 'snoRNA', 'snRNA' were excluded. Additionally, all Ensembl *miR-430* annotations were excluded, and a meta '*miR-430* hairpin' gene added in, based on the quantification described in the previous section. For intronic expression comparisons, because overall counts are lower, variance models for DESeq were calculated using both intronic counts and exonic counts as separate gene entries (that is, at most 1 intronic count entry and 1 exonic count entry per gene). Differential expression proceeded as normal, except multiple test correction of *P* values was applied relative only to the intronic counts.

Six sets of differential expression analyses were performed separately: exons and introns for each of (group 1) wild-type 64 cell, wild-type sphere, wild-type shield, U1U2 MO 4 h.p.f., α -amanitin 4 h.p.f. and α -amanitin 6 h.p.f., with the two α -amanitin conditions serving as pseudo replicates for DESeq for variance

estimation; (group 2) sphere stage wild type, Nanog MO, SoxB1 MO, Nanog MO plus SoxB1 MO, *MZpou5f1*, Nanog MO plus *MZpou5f1*, SoxB1 MO plus *MZpou5f1*, Nanog MO plus SoxB1 MO plus *MZpou5f1*, and two biological replicate shield stage wild-type samples for variance estimation; (group 3) shield stage wild-type, Nanog MO, two Nanog MO plus SoxB1 MO conditions treated as non-replicates, *MZpou5f1*, SoxB1 MO plus *MZpou5f1*, Nanog MO plus SoxB1 MO plus *MZpou5f1*, and two additional biological replicate shield stage wild-type samples to parallel group 2. For groups 2 and 3, we applied an exonic RPKM ≥ 1 and intronic RPKM ≥ 0.5 threshold in one or more of the samples.

Zygotic transcription was determined on the basis of significant exon and intron increases in sphere and shield stages relative to α -amanitin. 64 cell (pre-MZT) was used as further confirmation when no significant changes in intron level were detected or the gene was intronless (genes with <10 nt of unmasked intron sequence were considered effectively intronless). Increases in either exon signal, intron signal, or both determined positive zygotic transcription. For genes with a maternal contribution, increases in intronic signal due to zygotic transcription can be accompanied by no change or decreases in exonic signal. For genes significantly expressed, zygotic expression contribution is estimated using either intronic RPKM level or the RPKM difference between the post-MZT condition and the maximum of 64-cell and α -amanitin expression levels. Expression calls are provided in Supplementary Data 1.

To define first-wave genes, genes that were detected as transcribed in the U1U2 MO treated embryos above an expression level of 5 RPKM were considered to be first wave, using an estimate for zygotic transcription based on intronic signal for multi-exon genes, or comparison to α -amanitin and 64 cell for single-exon genes as described above. Although a cutoff of 5 RPKM was used for the main analyses, lower levels of transcription were observed for many genes, indicating weaker degrees of activation. Genes that were not called as transcribed in wild-type sphere were removed from the analysis.

Classification of loss-of-function expression categories. Significant changes in LOF conditions relative to wild type were determined using either intron or exon signal, depending on the pattern of signal originally used to call the gene as zygotically expressed. For genes with no maternal contribution, decreases in either exon or intron levels relative to wild type are considered to be loss of zygotic expression, whereas increases in either exon or intron levels are considered to be ectopic increases in zygotic expression. For genes with maternal contribution, we distinguish between two cases: (1) if zygotic transcription was originally detected in wild type only using intronic signal, then loss of zygotic transcription in the loss-of-function conditions is called only when intronic signal is lost; (2) if zygotic transcription was originally detected in wild type with both exonic and/or intronic signal, then decreases in either intronic levels or exonic levels indicate loss of zygotic expression, with intronic signal taking precedence when the directions of change disagree. For LOF embryos with the *MZpou5f1* genotype, differential expression was additionally performed between uninjected and injected *MZpou5f1* conditions, and expression differences between the injected conditions and wild type were required to be transitively consistent—for example, if a gene is called significantly lower in uninjected *MZpou5f1* than wild type, and a gene is significantly lower in injected *MZpou5f1* than uninjected *MZpou5f1*, then the gene must also be considered lower in the injected compared to wild type. To ensure that expression level differences in the *MZpou5f1* background are due to zygotic contributions, in addition to relying on intron signal, we filtered out any genes that were previously reported to be differentially maternally provided in *MZpou5f1* (ref. 19).

ChIP-seq analysis. Re-analysis of previously published Nanog ChIP-seq data (GSE34683) was performed as described²⁴, except using the current version of the zebrafish genome, Zv9. For *miR-430* locus alignment, reads were aligned exhaustively to the region chr4:27994413–28019085 (2 kb \pm the *miR-430* polycistron) using Bowtie with parameters -v 1 -best -strata -all. To estimate read depth and enrichment, reads were normalized by the number of times the read aligned the genome. To focus on the maximally non-redundant region in the locus, reads were preferentially aligned closest to the presumptive 5' boundary of the polycistron (chr4:28000732, corresponding to the 5' end of ENSDARG0000082539).

Morpholino oligonucleotide sequence. Sox2 MO1 5'-GAGAGGCTGCTGAA GTTACCTTAGC-3'; Sox2 MO2 5'-CTCGGTTTCCATCATGTTATACATT-3'; Sox3 MO1 5'-TACATTCTTAAAAGTGCTGCAAGC-3'; Sox3 MO2 5'-GAAG TCAGTCAAAAGTTCAGAGAGC-3'; Sox19a MO1 5'-GTACATGGCTGCCA ACAGAGTTAG-3'; Sox19a MO2 5'-AAAACGAGAGCGAGCCGTCTGTAA C-3'; Sox19b MO1 5'-GTACATCATGCCACTTCTCGCTTTG-3'; Sox19b MO2 5'-ACGAGCGAGCCTAATCAGGTCAAAC-3'; Nanog MO1 5'-CTGGCATCT TCCAGTCCGCCATTTC-3'; Nanog MO2 5'-AGTCCGCCATTTCGCCGTTA GATAA-3'; U1 MO1 5'-GGTATCTCCCTGCCAGGTAAGTAT-3'; U2 MO1 5'-TGATAAGAACAGATACACTTGA-3'; U2 MO2 5'-TATCAGATATT AAACGTATAAGAAC-3'.

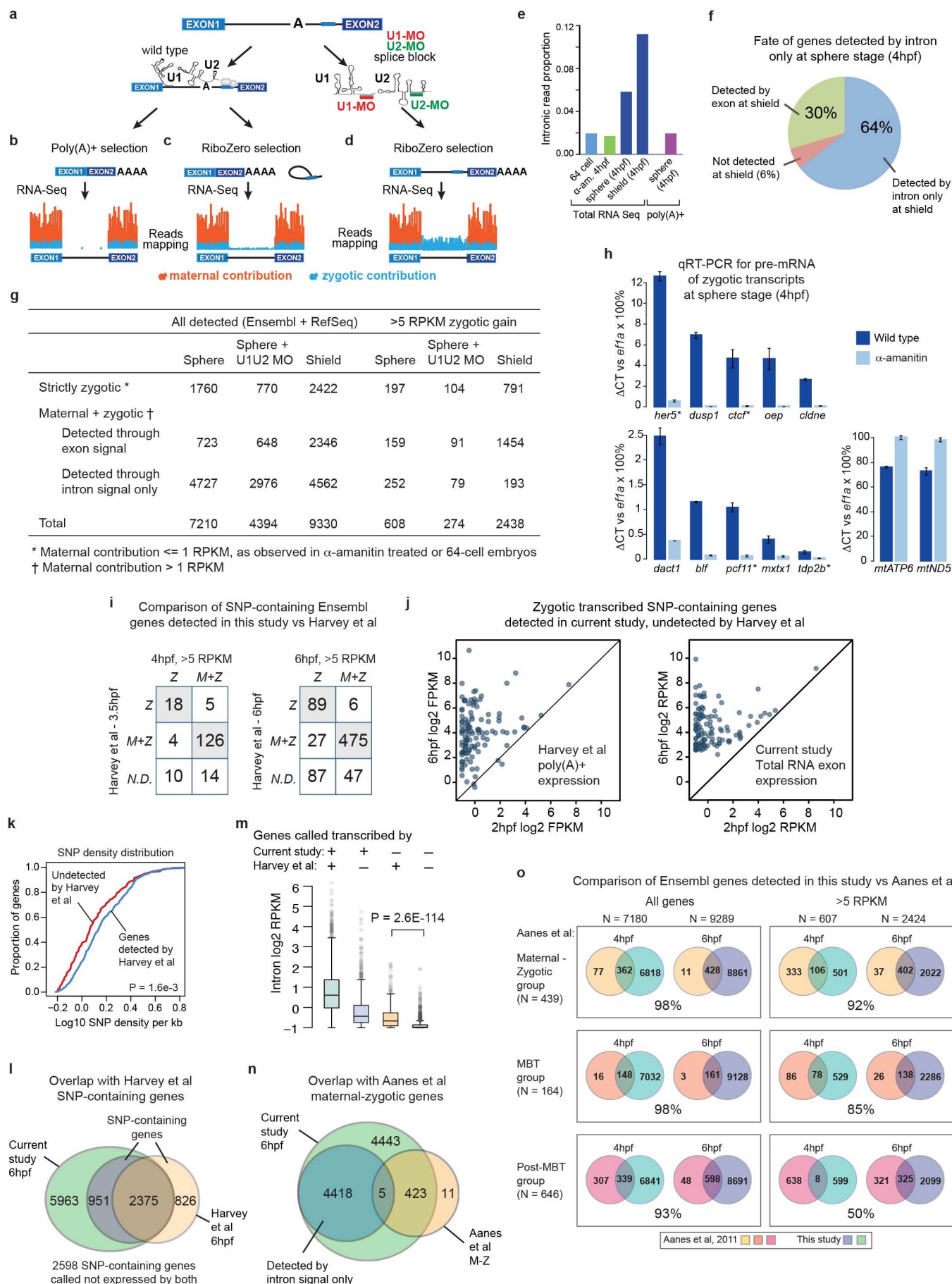
In situ primers. *ntl* forward 5'-TGGAAATACGTGAACGGTGA-3', reverse 5'-GTACGAACCCGAGGAGTGAA-3'; *isg15* forward 5'-AGAAGGGCCAGG

TCAAAACT-3', reverse 5'-*CATCACGGCATTGAAAACAC-3'; *cebpb* forward 5'-GTATGCAAGCAGCCAGTCAA-3', reverse 5'-*TGACTCGTCGCTGTCC TTG-3'; *cln* forward 5'-TGGTGTCTATGTGCCGAGAG-3', reverse 5'-*CGG CTGGGAGTATTTTCATGT-3'; *krt18* forward 5'-ATCACCGGCCTAAGAAAG GT-3', reverse 5'-*TCGTACTCTGCGTCTGATG-3'; *foxa3* forward 5'-CTTC AACGATTGCTTCGTCA-3' reverse 5'-*CATCTTCTGCTCGTTGGAC-3'; *vent* forward 5'-ACCCAGCAAGTTCTCAGTGG-3', reverse 5'-*TAGCAGCGTGTG AACAGCAT-3'; *nnr* forward 5'-CAGAGATGGACAGCGATTCA-3', reverse 5'-*TTCTGTTTCTTCTGCGGAGTTT-3'; *blf* forward 5'-GTCTCACAAGCGAATC CACA-3', reverse 5'-*GTGTGGGTCTTCTCGTGTT-3'. Asterisks indicate where a T7 promoter sequence gactTAATACGACTCACTATAGGG was added for *in vitro* transcription

RT-PCR primers. *nnr* forward 5'-AGCGTTTACAGCGGATCTCA-3', reverse 5'-*AGTGGACGGGGAAATAAAC-3'; *isg15* forward 5'-CGAAAGCCTCA TTCAGCAAC-3', reverse 5'-*GTGCAACTTCATGCCAGACTC-3'; *cln* forward 5'-TGGTGTCTATGTGCCGAGAG-3', reverse 5'-*CGGCTGGGAGTATTTT ATGT-3'; *sox11a* forward 5'-CGAAACGGACAGCATGTCTA-3', reverse 5'-GG AGTCGTCATCGTCGCTT-3'; *grhl3* (1/2) forward 5'-GAGGAGACCGGATA CCAAAC-3', reverse 5'-CCAAGCTCCACTGTGTTGT-3'; *grhl3* (1/3) forward 5'-GAGGAGACCGGATACCAAAC-3', reverse 5'-TTGTAAATGCTGCTCT CAGC-3'; *cln* forward 5'-ACTCCCATGTGGAAAGTCA-3', reverse 5'-GG GGTTCGCTGTATTAGC-3'; *krt4* forward 5'-GCAACCTCTCCACTCAC TC-3', reverse 5'-AATTGTGGGGTCAATTTCCA-3'; *hist1h2aa* forward 5'-CA AAGCTAAGACTCGCTCTCT-3', reverse 5'-TCTGTCTTCTTGGGACAGC-3'; *tubb4b* forward 5'-AGGTCTGGTCCATTGGTCA-3', reverse 5'-CATCCA GAACGGAATCAACC-3'; *klf4b* forward 5'-ACAGTTGTGAATTCCTGGAT G-3', reverse 5'-GTTTACATGTGCCTCTTCATGTG-3'; *vox* forward 5'-GAC TGGCTTGCTCAGAGCTT-3', reverse 5'-GGCCGCTTCACTCTCATAAC-3'; *tbx16* forward 5'-AACCTTACCTTCCCGAGA-3', reverse 5'-CAAGACTCG GGAACAAAGC-3'.

qRT-PCR primers. *blf* forward 5'-CCCTGCTGAGCTTGCATAGT-3', reverse 5'-CCCACACTGAGGACACTTGA-3'; *cln* forward 5'-GGCTTCTTGGGAG CCATTAT-3', reverse 5'-GCGAAAAAGCTGACGATGAT-3'; *ctcf* forward 5'-GTTAGCAGAGGCTTGCTTTACTG-3', reverse 5'-GCAGTGAATTTCCGC ACA-3'; *dact1* forward 5'-AGCCTCGGTTCTTCTTCAACA-3', reverse 5'-GGA GGATTTGTGCAAGTGGT-3'; *dup1* forward 5'-CTCCAGTAATGTGCGCTT CA-3', reverse 5'-TGGTCGAACCTTTTACCTTCA-3'; *efla* forward 5'-TGAT CTACAAATGCGGTGGA-3', reverse 5'-CAATGGTGATACCACGCTCA-3'; *her5* forward 5'-CCAAGCCTCTCATGGAGAAA-3', reverse 5'-TAGCTCTGA CGTTTGATGG-3'; *mtATP6* forward 5'-CTTAGCGGCCACAATAGAG-3', reverse 5'-ATGGGGGTTCTTCTGGTAA-3'; *mtND5* forward 5'-TTCTTAT GCTCAGGGGCAAT-3', reverse 5'-TTAGGGCTCAGGCGTTAAGA-3'; *mxtx1* forward 5'-GAAATGCAAGGTTGGA-3', reverse 5'-ACCCAGTTAGG AGGCATCT-3'; *oep* forward 5'-TTCTGGAAAGCCAAAGCAAT-3', reverse 5'-TCATGTGCTGTGCGAGTTG-3'; *pcf11* forward 5'-CCTCGCTGGAAGATC TGACT-3', reverse 5'-CATGTTACAGGCTCATGTCA-3'; *tdp2b* forward 5'-GG AGCCACCTGCTCTATTA-3', reverse 5'-ACCTGCCAATTGTGAAGATA-3'.

48. Amsterdam, A. et al. Identification of 315 genes essential for early zebrafish development. *Proc. Natl Acad. Sci. USA* **101**, 12792–12797 (2004).
49. Matter, N. & König, H. Targeted 'knockdown' of spliceosome function in mammalian cells. *Nucleic Acids Res.* **33**, e41 (2005).
50. Rösel, T. D. et al. RNA-Seq analysis in mutant zebrafish reveals role of U1C protein in alternative splicing regulation. *EMBO J.* **30**, 1965–1976 (2011).
51. Iwafuchi-Doi, M. et al. The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates. *Dev. Biol.* **352**, 354–366 (2011).
52. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995).
53. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
54. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
55. Collins, J. E., White, S., Searle, S. M. & Stemple, D. L. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res.* **22**, 2067–2078 (2012).
56. Berg, M. G. et al. U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53–64 (2012).
57. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
58. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
59. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6**, 117–131 (2004).



Extended Data Figure 1 | Identifying *de novo* zygotic transcription.

a, Schematic of the sequencing strategy used in this study. Most zebrafish protein-coding genes (>95%) contain introns. *De novo* transcription produces intronic RNA sequences, which are spliced out of pre-mRNAs by the spliceosome, consisting of several ncRNA species including U1 and U2.

b, Typical mRNA-seq applications use poly(A)⁺ selection to enrich for the mature mRNA population. Sequence reads map predominantly to exonic regions, with very few reads mapping to introns. During embryogenesis, many zygotic transcribed genes are expected to have a maternal contribution in the cytoplasm from the oocyte. The resulting signal will be a mixture of maternal-derived (orange) and zygotic-derived (blue) mRNA molecules, which cannot be deconvoluted without comparing to a reference sample to look for exon expression level change.

c, mRNA-seq applications that skip poly(A)⁺ selection and instead use a rRNA depletion protocol (RiboZero) will not enrich for the mature mRNA population. Thus, transcripts in all stages of biogenesis (pre-mRNA, partially spliced mRNA, spliced introns) will be sequenced, and reads are expected to map both to exons and introns. Because maternally contributed mRNAs are mature, any intron signal detected must derive from *de novo* zygotic transcription. To determine the background signal for each intron, α -amanitin is used as a negative control for transcription.

d, Morpholinos complementary to U1 and U2 injected into one-cell embryos inhibit zygotic splicing. Thus, pre-mRNAs fail to be processed, and the entire population of zygotic mRNAs will be unspliced. There are two benefits: (1) intron signal is amplified, as introns are stabilized in the pre-mRNA compared to spliced out introns; (2) protein production from zygotic mRNAs is effectively halted, as pre-mRNAs are generally not competent for normal translation. Only the first wave of transcription, resulting from activation by maternal factors, is observed. Transcription that requires zygotic proteins (subsequent waves) will be largely absent.

e, The proportion of sequencing reads aligning to gene introns. Total RNA sequencing reveals elevated intronic sequence reads, corresponding to *de novo* zygotic transcription.

f, The fate of the 5,318 sphere-stage (4 h.p.f.) zygotic genes that are only detectable through significant changes in intron sequence. At shield stage (6 h.p.f.), 64% of the genes are still detected as zygotically transcribed based only on intron signal. These include genes that have simultaneous zygotic transcription with decay of the maternal contribution. 30% of the genes are detected using both exon and intron signal by shield stage, indicating that transcription levels at sphere stage were too low to detect differences in exons, but were apparent in the introns.

g, Number of genes detected in wild-type sphere-stage embryos, sphere embryos injected with U1U2 MO and wild-type shield-stage embryos, at different thresholds of detection. For both groups, a multiple test-corrected $P < 0.1$ threshold (Benjamin–Hochberg) was used for differential expression of exonic signal. For intronic signal, an uncorrected $P < 0.1$ was used for the ‘All detected’ group, whereas a multiple test-corrected $P < 0.1$ was used for the >5 RPKM gain group.

h, Quantitative RT-PCR was performed for select genes to confirm zygotic transcription in wild-type sphere-stage embryos (dark blue bars) compared to α -amanitin-treated embryos (light blue bars). Primers were designed to amplify pre-mRNAs across exon–intron boundaries, except for *cldn*. Expression levels are reported as percentage of CT value compared to a maternally provided housekeeping gene (*ef1a*) ($\Delta CT \times 100\%$). Error bars show

s.e.m. for three technical replicates. Increased pre-mRNA levels were observed for all zygotic genes tested between wild type and α -amanitin. Maternally provided genes *mtATP6* and *mtND5* show no increase in wild type. Genes marked with an asterisk represent the bottom 10% of significant differential intron expression based on the RNA-seq data (which quantifies both pre-mRNA and spliced introns). This shows that using intron signal is a reliable indication of zygotic transcription.

i, Genes detected in this study were compared to previous annotations of zygotic transcripts¹³, which used SNPs to identify transcripts derived from paternal alleles, to distinguish zygotic transcription from the maternal contribution. From their genomic sequencing results, we extracted 6,750 genes with informative exonic SNPs, which were consistently called between the two sets of matings. 178 of the genes we call zygotically transcribed at sphere stage at levels >5 RPKM are among the 6,750 informative genes. 87% of these are also found to be transcribed by ref. 13, with agreement between both strictly zygotic genes (Z) and maternal+zygotic genes (M+Z). 24 genes were not detected by ref. 13 (N.D.). At shield stage, 82% of the zygotic genes are also found by ref. 13, with 134 genes not detected.

j, These undetected genes nevertheless have highly increased expression pre-64-cell to post-MZT (shield) using the RNA-seq data generated by ref. 13 (left) and in the current study (right).

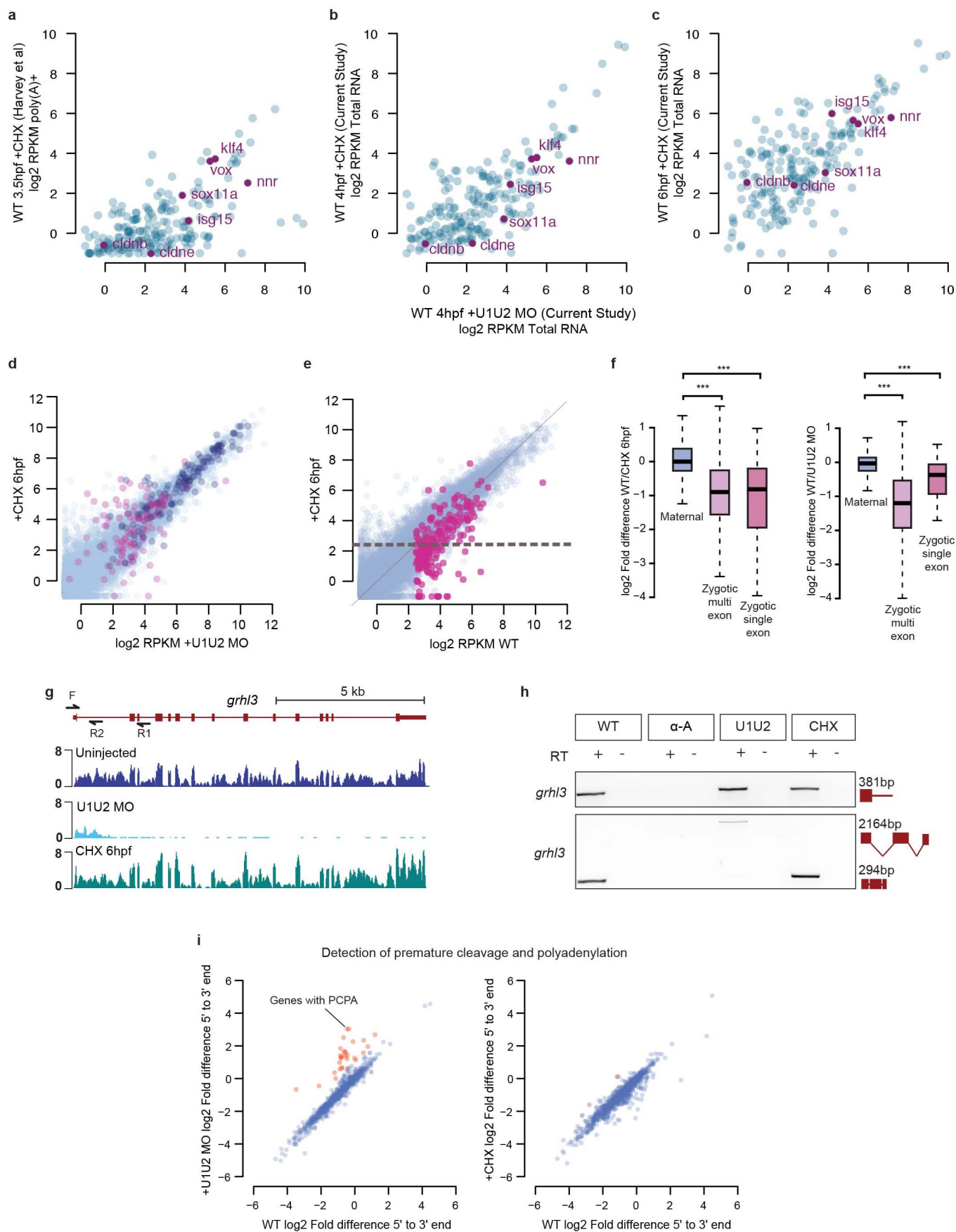
k, Cumulative plots show that SNP density is significantly lower among ref. 13 undetected genes at shield compared to detected genes ($P = 1.6 \times 10^{-3}$, two-sided Wilcoxon rank sum test), suggesting that low SNP density may account for the missed genes.

l, Overall, ref. 13 and the current study distinguish a similar number of zygotic versus maternal transcripts at 6 h.p.f., among Ensembl genes with informative SNPs, with 74% agreement. However, 64% of zygotic transcripts identified in the current study do not have informative SNPs, and are thus not called transcribed by ref. 13.

m, Genes called transcribed by ref. 13 but not in the current study have significantly higher intron signal than maternal genes ($P = 1.4 \times 10^{-95}$, two-sided Wilcoxon rank sum test), indicating that our significance threshold to detect zygotic transcription is conservative.

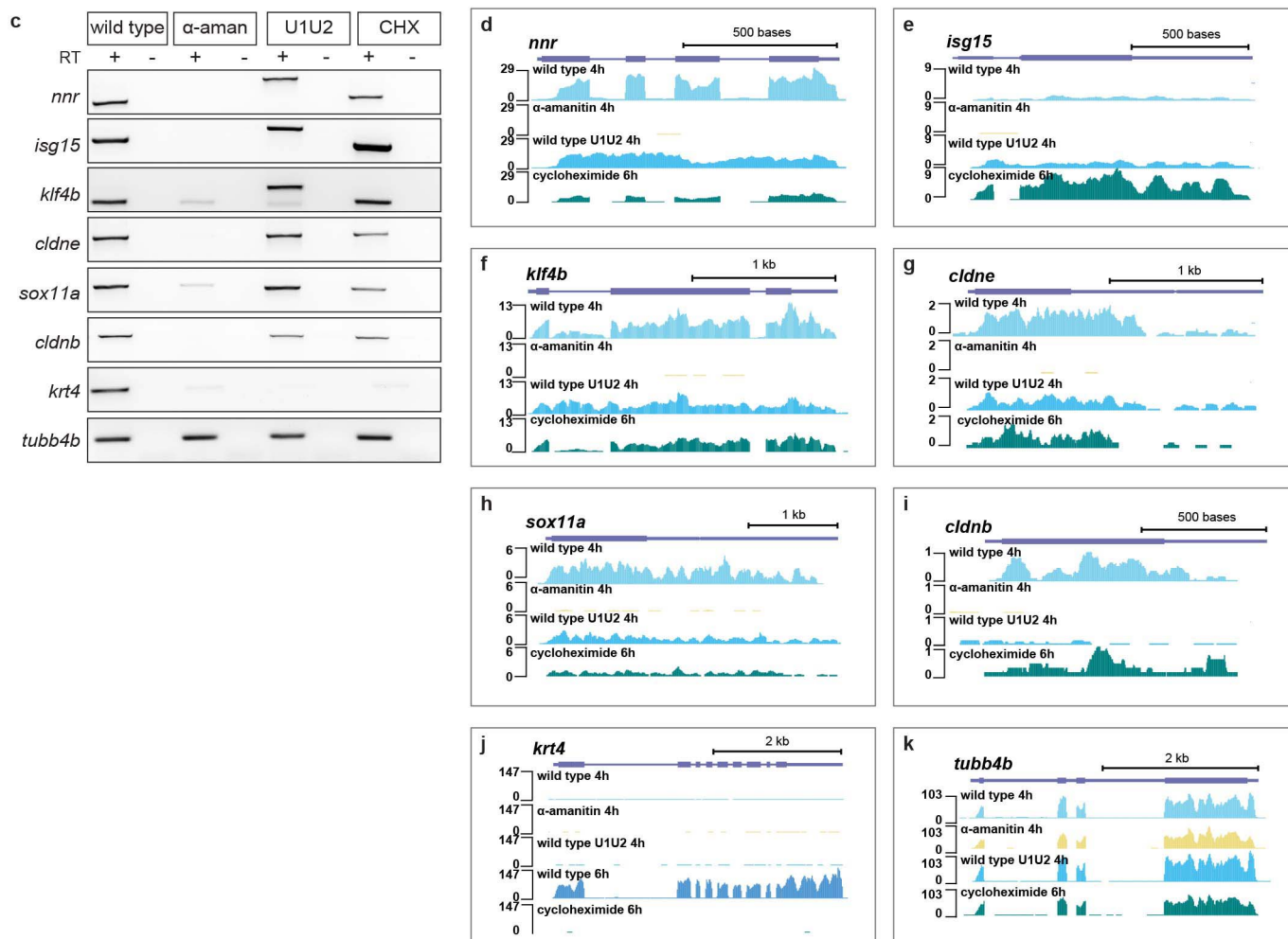
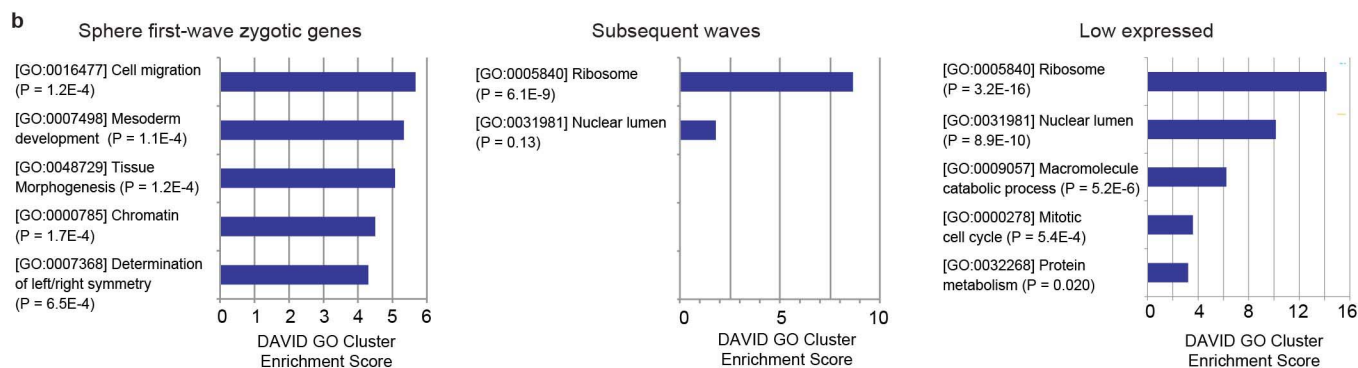
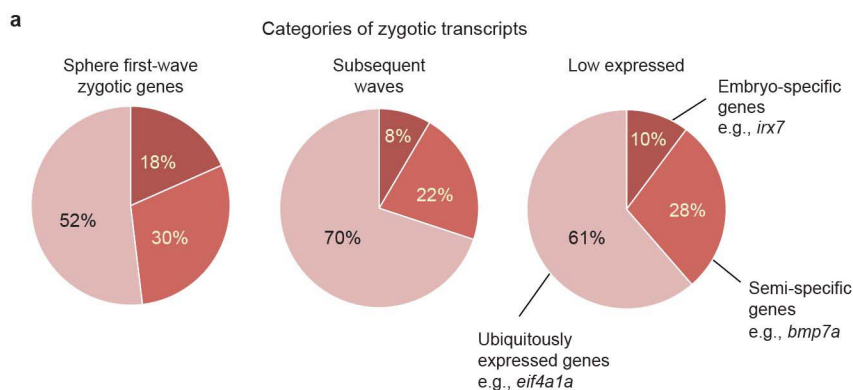
n, Reference 14 used a time course poly(A)⁺ RNA-seq strategy to define zygotic transcripts. The comparable r70 Ensembl genes in the ref. 14 maternal+zygotic gene category are largely found in our study; however, we find thousands more transcribed genes based on intron signal—these genes represent transcription that is masked by the maternal contribution.

o, Overall, our study captures most of the zygotic genes in the three categories described by ref. 14: maternal-zygotic genes (zygotic genes with maternal contribution, yellow), MBT genes (strictly zygotic genes detected at MBT, 3.5 h.p.f., orange), and post-MBT genes (strictly zygotic genes detected at 5.3 h.p.f., pink). Venn diagrams show the number of comparable r70 Ensembl genes that overlap between the two studies. Left panels include all zygotic genes detected in this study; right panels impose a zygotic expression threshold of >5 RPKM. Percentages within each box are calculated as the number of genes detected in this study (at either time point) that overlap the respective ref. 14 group, divided by the size of the ref. 14 group. The overlap percentages are generally high, indicating that our study recovered genes previously annotated as zygotically transcribed as well as many additional zygotic genes based on the use of intronic reads.



Extended Data Figure 2 | Cycloheximide and U1U2 MO transcriptomes show first-wave genes. **a–c**, Biplots comparing strictly zygotic genes found by either the current study or ref. 13 at >5 RPKM ($N = 202$). Zygotic expressed genes of ref. 13 were identified by comparing their raw RNA-seq data at 128-cell (pre-MZT) versus 3.5 h.p.f. In **a**, zygotic expression in U1U2 MO treated embryos (Total RNA, 4hpf) is compared to ref. 13 embryos treated with cycloheximide (CHX) (poly(A)⁺, assayed at 3.5 h.p.f.), which shows lagging expression of many first-wave genes (defined as having >5 RPKM in U1U2 MO). Genes verified by RT-PCR as first wave (*klf4*, *nnr*, *sox11a*, *isg15*, *clde*) are highlighted, in addition to *clde*, which misses the threshold for first wave in the U1U2 MO transcriptome, and *vox*, which was highlighted by ref. 13. In **b**, **c**, Embryos treated with CHX and assayed in the current study at 4 h.p.f. and 6 h.p.f. (Total RNA) show gradual increases in expression of zygotic genes. Together these results suggest that expression of first-wave genes is independent of *de novo* zygotic factors, and that transcription overall is slower in CHX-treated embryos compared to wild type or U1U2 MO. **d**, Biplot showing gene expression levels (exonic) for all genes in U1U2 MO embryos at 4 h.p.f. compared to CHX-treated embryos assayed at 6 h.p.f. Magenta points, strictly zygotic genes; dark-blue points, maternal + zygotic genes. 97% of the first-wave genes called in U1U2 MO were expressed >1 RPKM in the CHX condition. **e**, Biplot comparing exonic expression levels between wild-type (4 h.p.f.) and CHX-treated embryos. Magenta points are strictly zygotic genes expressed >5 RPKM in wild type. The dotted line indicates 5 RPKM expression in CHX. **f**, Box-and-whisker plots comparing exonic expression level differences between wild-type and treated embryos in maternal genes, strictly zygotic multi-exon genes, and strictly zygotic single-exon genes. Both U1U2 MO and CHX-treated embryos show loss of expression in zygotic genes compared to wild type (U1U2 MO: $P = 9.4 \times 10^{-207}$ for multi-exonic, $P = 4.2 \times 10^{-4}$ for single exon, Wilcoxon rank-sum test comparing to maternal; CHX: $P = 4.3 \times 10^{-137}$ multi-exon, $P = 1.5 \times 10^{-6}$ single exon). The box defines the first and third quartiles, with the median indicated with a thick black line. The systemic decreases in expression in the U1U2 MO or CHX conditions compared to wild type indicate that although maternal factors can

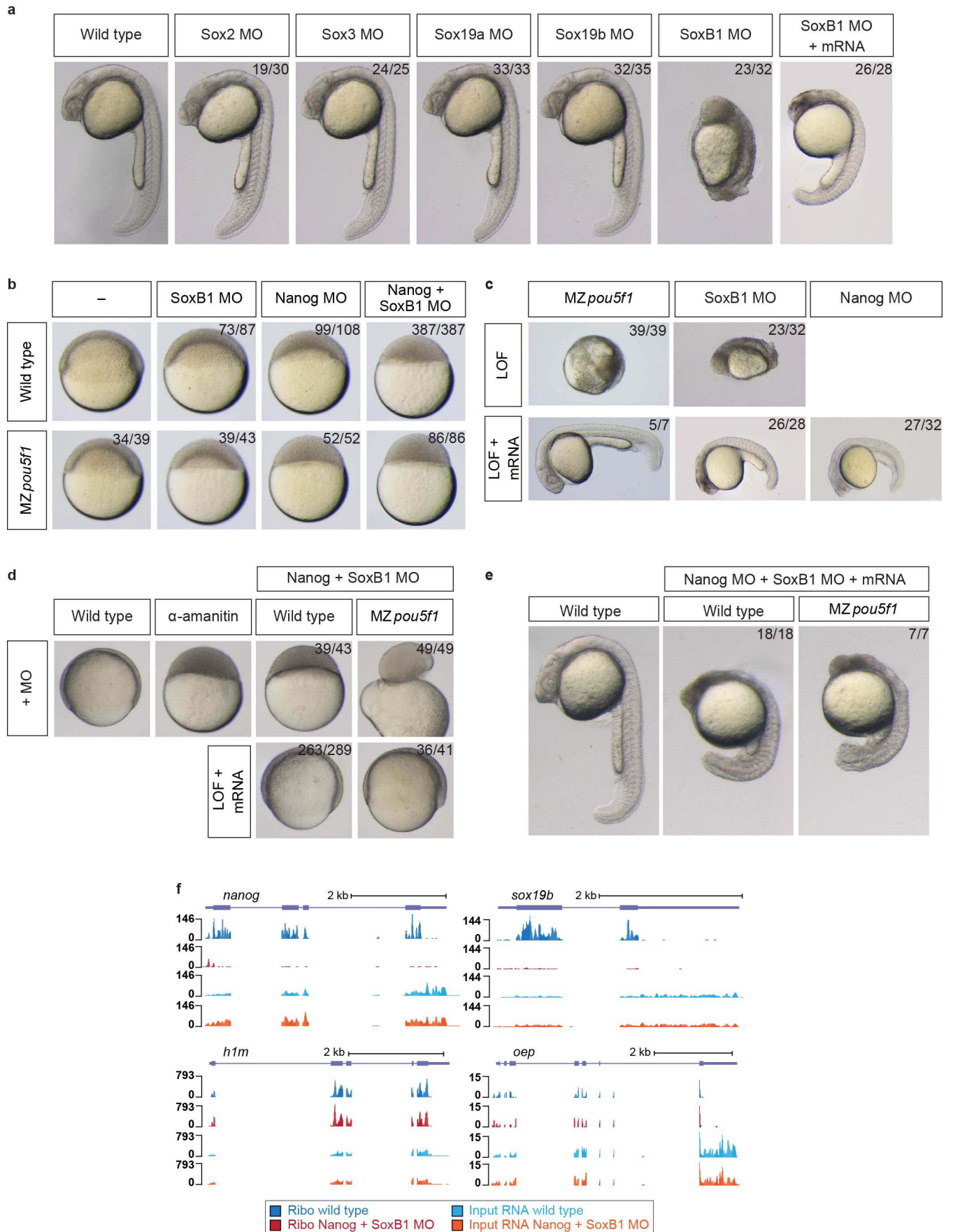
activate to a large extent expression of the first-wave genes, additional zygotic contribution of transcription factors (Nanog, SoxB1 and Pou5f1, but possibly others as well) might be required to reach wild-type levels of expression for many genes. This was also observed in ref. 13 for the gene *vox*. Alternatively, lower expression of first-wave zygotic genes might be caused by reduced level of maternal encoded proteins, as incubation with CHX at 32-cell stage might also decrease translation of the maternally deposited mRNAs. We consistently observe that CHX-treated embryos show lower/delayed expression compared with U1U2-MO-treated embryos, indicating that premature inhibition of maternal mRNA translation has an effect on the rate of activation of the first-wave genes. **g**, UCSC Genome Browser track showing an example of premature cleavage and polyadenylation (PCPA) for *grhl3*. Arrows indicate primer sites for RT-PCR. Previously, it was shown that U1 snRNA also serves to protect nascent mRNAs from PCPA, and that U1 inhibition results in 3'-truncation that may affect transcript level quantification⁵⁶. **h**, RT-PCR for *grhl3* on shield-stage embryos ($N = 5$). Wild-type (WT), U1U2 MO and CHX-treated embryos all amplify a 381-bp fragment from exon 1 to the beginning of intron 1. U1U2-MO-injected embryos amplify an unspliced 2,164-bp gene product spanning exon 1 to 3, whereas wild-type and CHX-treated embryos have a 294-bp spliced product, with α -amanitin as a negative control. **i**, Biplots comparing expression levels at the 5' end of a transcript compared to the 3' end, to detect PCPA at 4 h.p.f. Read density was assayed in up to 1,000 nucleotides of 5' and 3' sequence per transcript. The range of asymmetry values in wild type reflects sequencing biases or transcript annotation irregularities. Several genes in U1U2 MO embryos show elevated asymmetry compared to wild-type (orange dots, >twofold), reflecting a drop-off of read density moving 5'–3' in the transcript, indicative of PCPA. These genes are included in our annotations of the zygotic first wave of expressed genes. The minor extent of PCPA during embryogenesis may reflect the short length of many of the zygotic genes, as PCPA is associated with longer genes that are likely to harbour cryptic polyadenylation sites. Transcripts in CHX-treated embryos generally do not show this trend.



Extended Data Figure 3 | Verification of first-wave gene expression and functional categories.

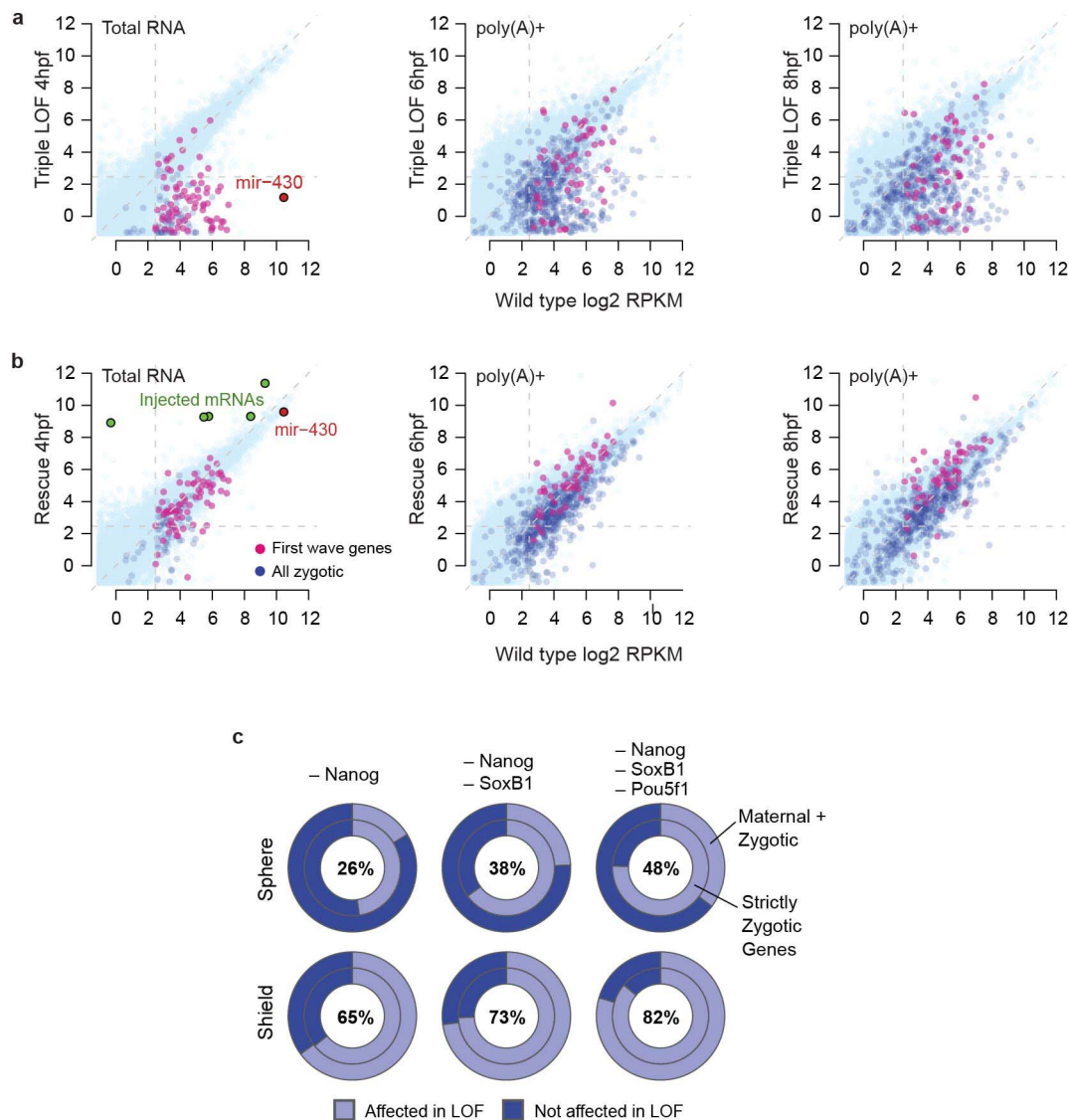
a. To assay the embryonic specificity of the first-wave genes, we used publicly available microarray data from NCBI GEO across eight normal adult tissue types (brain, GSE11107; liver, GSE11107; heart, GSE17993; skin, GSE24528; kidney, GSE32363; digestive tract, GSE35889; ovary, GSE14979; testis, GSE14979) to classify genes as expressed specifically in the embryo (called 'present' by the MAS5 algorithm in 0–2 different adult tissues), genes expressed semi-specifically (present in 3–5 different adult tissues), and genes expressed ubiquitously (present in 6–8 different adult tissues); this latter group would correspond to 'housekeeping' genes. Sphere-stage first-wave genes consist of a mixture of specifically expressed and housekeeping genes. Subsequent-wave genes and genes expressed at levels <5 RPKM consist of a larger proportion of genes typically expressed ubiquitously in adult fish, suggesting a widespread activation of genes encoding general cellular processes in addition to developmentally specific ones. **b.** Gene Ontology enrichment analysis for first-wave, subsequent-wave and the low expressed genes with intronic RPKM >0.5. Top 5 scoring clusters are shown for each gene set. Clusters were defined using DAVID (<http://david.abcc.ncifcrf.gov>) Gene

Functional Annotation Clustering on GO 'FAT' annotations and 'high' stringency. Clusters are annotated with representative GO terms and corresponding Benjamini–Hochberg FDR corrected *P* values. **c.** To validate genes activated in the first wave versus subsequent waves, RT–PCR was performed on shield stage (6 h.p.f.) in wild-type, α -amanitin, U1U2 MO and cycloheximide (CHX)-treated embryos. The unspliced products for *nnr*, *isg15* and *klf4* are detected only in U1U2 morphants, confirming that U1U2 is indeed blocking splicing. CHX treatment indicates the single-exon genes *cldne* and *sox11a* are activated in the first wave. *cldnb* is detected at low levels in wild type, as well as both U1U2 MO and CHX-treated embryos; however, based on RNA-seq levels at sphere stage, this gene does not pass the expression threshold to be called first wave. *krt4* is significantly reduced in U1U2 MO and CHX-treated embryos, indicating that zygotic factors are required for its activation. Maternal *tubb4b* is present in all conditions. **d–h.** UCSC Genome Browser tracks for first-wave genes *nnr*, *isg15*, *klf4*, *cldne* and *sox11a*. **i.** UCSC Genome Browser track for *cldnb*, which shows low expression levels at sphere stage. **j, k.** UCSC Genome Browser track for a gene activated in subsequent waves (*krt4*) and for a maternally provided gene (*tubb4b*).



Extended Data Figure 4 | Loss-of-function and rescue for Nanog, SoxB1 and Pou5f1. **a**, Wild-type embryos were injected with Sox2, Sox3, Sox19a and Sox19b morpholinos individually and in combination (0.125 mM). Consistent with other reports, only quadruple LOF results in severe developmental defects (27 h.p.f.)²⁰. LOF phenotype is rescued by injecting *soxB1* mRNA (imaged at 24 h.p.f.). **b**, Wild-type and MZ*pou5f1* embryos were injected with *SoxB1* MO (0.125mM each) and *Nanog* MO (0.6mM each) individually and in combination (*Nanog* + *SoxB1*). Loss of *Nanog* results in severe gastrulation defects and failure to progress past 80% epiboly, as previously reported²⁴. Loss of *SoxB1* in both wild-type and MZ*pou5f1* embryos showed developmental delay, whereas combined LOF for *Nanog*/*SoxB1* or *Pou5f1*/*Nanog* completely arrested development before epiboly. Triple LOF embryos also arrested and failed to undergo gastrulation. **c**, Individual LOF for *Nanog*, *SoxB1* and *Pou5f1* resulted in developmental abnormalities (top panel). Embryos with *Nanog* LOF did not progress past 80% epiboly. The LOF phenotypes were rescued by injecting the respective mRNAs (LOF + mRNA) (bottom panel). Embryos

imaged at 23 h.p.f. **d, e**, Wild-type and MZ*pou5f1* embryos were co-injected with *Nanog* + *SoxB1* MO. LOF embryos arrest at sphere stage and resemble α -amanitin-injected embryos (+MO). Combinatorial LOF is rescued with co-injection of the respective mRNAs (MO + mRNA). Embryos were imaged when wild-type siblings reached 80% epiboly (**d**) and 24 h.p.f. (**e**). **f**, Ribosome profiling was performed at 2 h.p.f. on wild-type embryos and embryos injected with *Nanog* and *SoxB1* morpholino at one-cell stage, to determine the specificity of the morpholinos to repress translation of *nanog* and *soxB1* mRNA. Sequenced ribosome protected fragments (RPFs) were predominantly 28–29 nucleotides long, indicative of the width of the ribosome footprint. UCSC Genome Browser tracks (sense strand) showing ribosome profiling (top 2 tracks per gene) and input mRNA (bottom 2 tracks per gene). *nanog* and *sox19b* show significant reduction in RPFs in the *Nanog* MO + *SoxB1* MO injected embryos compared to wild type. Input mRNA is unaffected. Neither *h1m*, a highly expressed gene, nor *oep*, a low expressed gene, has any change in either RPFs or input mRNA between wild-type and injected embryos.



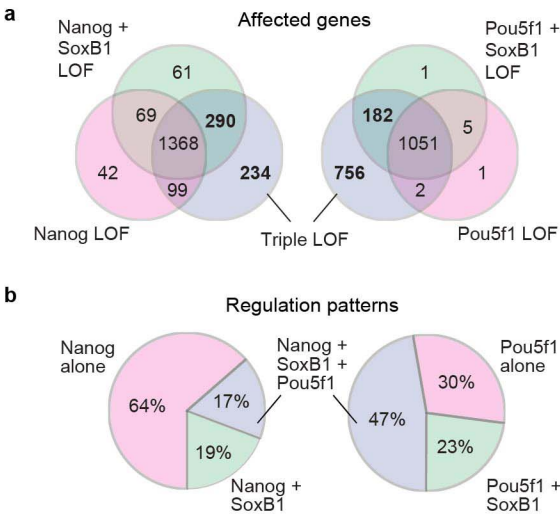
Extended Data Figure 5 | A transcriptome-wide effect is observed in LOF embryos. **a, b,** Biplots comparing log₂ RPKM exonic expression levels between time-matched wild-type and Nanog + SoxB1 + Pou5f1 LOF embryos (**a**); and between wild-type and triple LOF embryos co-injected with mRNA for *nanog*, *soxB1* and *pou5f1* (**b**) at 4 h.p.f., 6 h.p.f. and 8 h.p.f. Dark blue points highlight all strictly zygotic genes, whereas magenta points highlight the first-wave zygotic genes. miR-430 is highlighted at 4 h.p.f. in red, whereas green points indicate expression levels of (left to right) *sox2*, *sox3*, *sox19a*, *sox19b* and *nanog*. **c,** Plots showing proportion of the zygotic transcriptome affected (including first and subsequent waves). For sphere and shield stages and each

LOF (Nanog MO, Nanog MO + SoxB1 MO, MZ*pou5f1* + Nanog MO + SoxB1 MO), dark blue regions represent genes with normal expression compared to wild type; light blue regions represent genes with significant loss of expression. Inner ring comprises zygotic genes with <1 RPKM of maternal contribution; outer ring comprises zygotic genes with maternal contribution. Percentages represent total affected genes in that condition over both gene categories. At sphere stage (4 h.p.f.) the effect for maternal and zygotic (M+Z) genes is weaker than for strictly zygotic genes, which may reflect a reduced power to detect changes due to the maternal contribution (see also Fig. 3b).



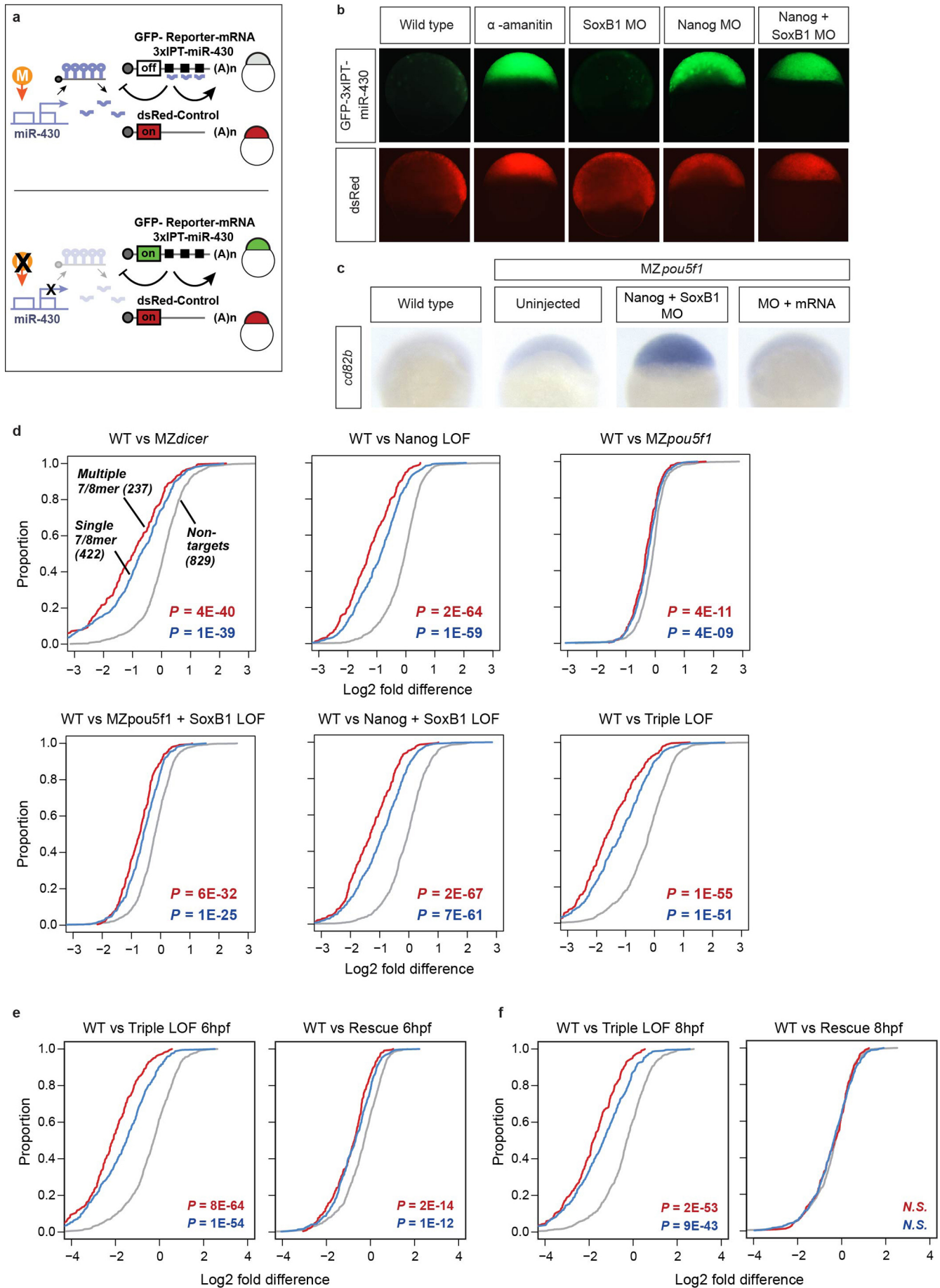
Extended Data Figure 6 | Zygotic genes fail to be activated with Nanog, SoxB1 and Pou5f1 LOF. **a–f**, *In situ* images showing that loss of Nanog and SoxB1 function results in a significant reduction in zygotic *foxa3*, *blf*, *vent*, *foxd3*, *krt18* and *ntla* expression. LOF embryos (Nanog + SoxB1 MO) resemble α -amanitin-injected embryos by *in situ*, as well as in their transcriptome profiles. Loss of Nanog and SoxB1 is rescued by *nanog* and *soxb1* mRNA (MO + mRNA), which is sufficient to restore wild-type expression profiles. **g, h**, *In situ* hybridization for zygotically transcribed *cldne* and *cebpb* shows that loss of Nanog and SoxB1 (Nanog + SoxB1 MO) has minimal effect on

activation of *cldne* and *cebpb*. However, triple LOF shows a decrease in expression for both genes, as shown in the UCSC tracks. **i–o**, RT-PCR analysis (**i**) and UCSC Genome Browser tracks (**j–o**) for zygotic genes *klf4b*, *vox*, *tbx16*, *mxtx2*, *her3* and *sox32*, showing differential expression of zygotic genes in LOF conditions. Expression levels were rescued by injecting *nanog* and *soxb1* mRNA (MO + mRNA). Maternal *hist1h2aa* was present in the α -amanitin control. RT (–) indicates the absence of reverse transcriptase, to control for genomic DNA contamination. In UCSC tracks, loss of Nanog, SoxB1 and Pou5f1 in each sequenced condition is indicated by (–).



Extended Data Figure 7 | Loss of function affects genes across functional categories in a combinatorial manner. **a**, Comparisons of the single and double LOF transcriptomes to the triple LOF reveal that regulation is often combinatorial and redundant. Although all three factors seem to exert some influence on most of the transcribed genes, the effects observed in the combined LOF are not usually additive. Nanog seems to have the strongest individual effect of the three factors, but Pou5f1/SoxB1 can often act redundantly, or amplify the effect of Nanog alone. Venn diagrams show overlap between genes significantly downregulated at shield stage in single (pink), double (green) and triple (blue) LOF embryos. $n = 2,172$, left; $n = 2,027$, right. **b**, Pie charts showing the relative influence of each factor in the triple LOF. For each pie chart, genes downregulated in the triple LOF were compared in the single and double LOF transcriptomes. If the downregulation of a gene observed in the single LOF was less than twofold different from that observed in the triple LOF, the gene was considered to be regulated by the single factor alone. Otherwise, if the downregulation in the double was less than twofold different than the triple LOF, the gene was considered regulated by the combination of two factors. All remaining genes display the strongest downregulation in the triple LOF. Note that genes in each category may be affected by other combinations of LOF; however, the effect there is weaker. **c**, Breakdown of effects showing the redundancy of regulation in genes downregulated in the triple LOF. The largest category of genes seems to be regulated exclusively by Nanog (31%), as loss of Nanog function is equivalent to the triple LOF. 16% of genes seem to be regulated by both Nanog and Pou5f1 together, as loss of either Nanog alone or

loss of Pou5f1 alone is sufficient to achieve the loss of function observed in the triple LOF. 16% of genes have equivalent effects with either Nanog LOF or Pou5f1 + SoxB1 double LOF, suggesting that Pou5f1 and SoxB1 act redundantly for these genes to co-regulate with Nanog. 9% of genes show the strongest effect only in the triple LOF. This suggests that there is redundancy between all three factors, as these genes can still be activated when one or two factors are lost. In all, 76% of the affected genes are subject to some form of redundant or combinatorial regulation. Asterisk indicates that for genes where the effect in the triple LOF was equivalent to both the double loss of SoxB1 and Nanog, and the double loss of SoxB1 and Pou5f1, we inferred that the effect was conferred by SoxB1 alone. **d**, Most genes are affected in the double or triple LOF conditions, across the gene categories defined in Extended Data Fig. 3a, including both embryo-specific genes and housekeeping (ubiquitously expressed) genes. **e**, Heat map showing specific embryonic functional categories of genes downregulated in LOF embryos. Three GO categories of genes expressed in wild type at shield stage are shown: general transcription factors, gastrulation and cell movement genes, and patterning genes (anterior–posterior axis and dorsal–ventral axis). Expression levels are represented as row-normalized values on a red–green colour scale for wild type (WT), α -amanitin treated (A), Nanog LOF (N), Nanog + SoxB1 LOF (NS), and Nanog + SoxB1 + Pou5f1 triple LOF (NSP). Widespread loss of expression is observed across these functional categories, with the triple LOF exhibiting the greatest similarity to α -amanitin.



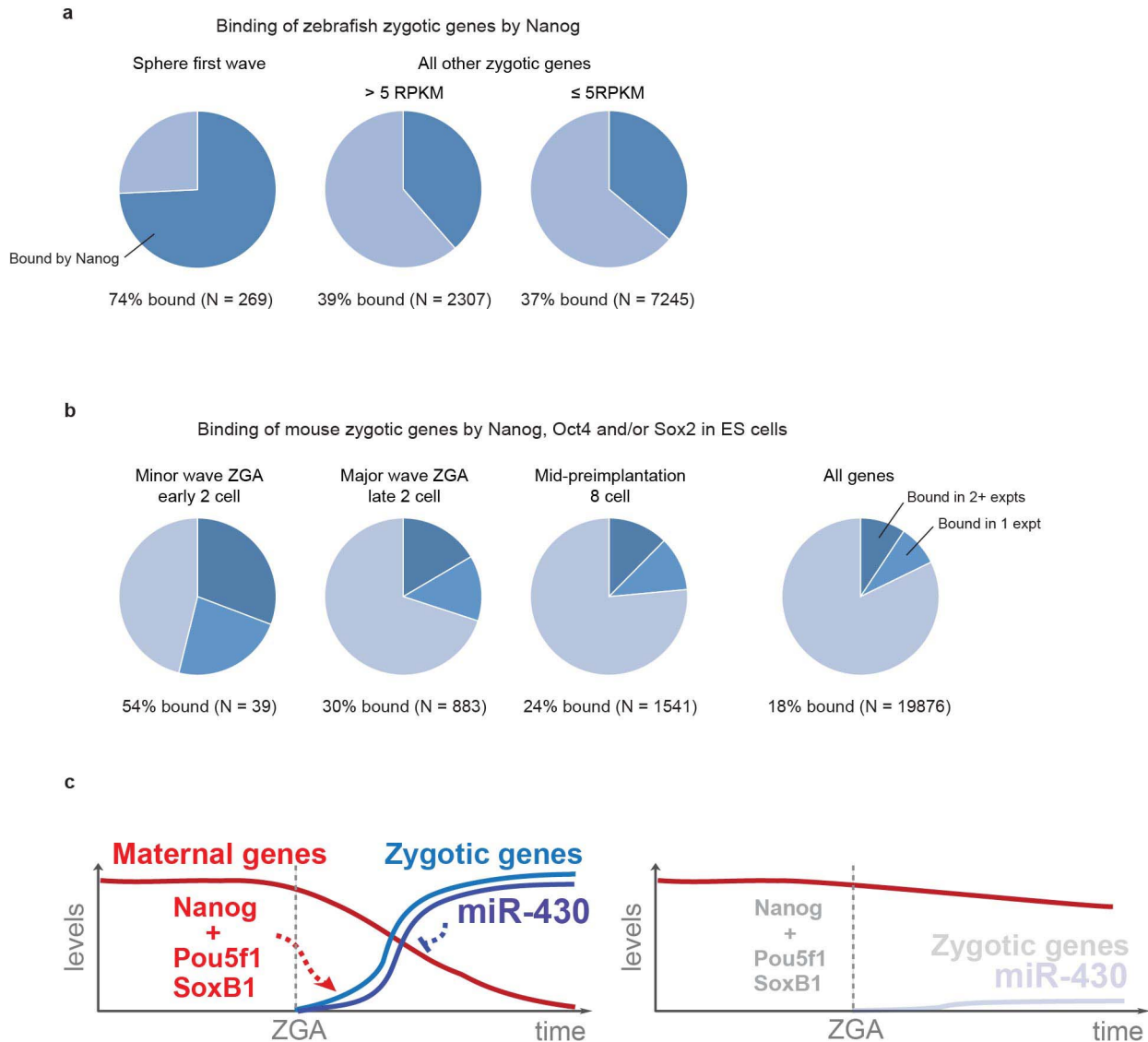
Extended Data Figure 8 | miR-430 activity requires Nanog function.

a, Schematic representation of miR-430 activity reporter GFP-3×IPT-miR-430 containing three complementary target sites to miR-430 (ref. 26). If maternal factor (M) is present, miR-430 is expressed and represses translation of the target mRNAs (no GFP expressed). Conversely, loss (X) of the maternal factor required for miR-430 activation would lead to a failure to repress miR-430 targets and GFP expression. dsRed is a control mRNA that is not subject to regulation by miR-430 and is co-injected with the target mRNA.

b, GFP-reporter and dsRed (injection control) mRNAs were co-injected into embryos at one-cell stage and fluorescence assayed 7–8 h.p.f. GFP-reporter is repressed in wild-type and SoxB1 morphants by endogenous miR-430 (ref. 26), as shown by a decrease in GFP expression. The GFP-reporter fails to be repressed in α -amanitin (that fail to activate zygotic transcription and do not express miR-430) and Nanog-MO-injected embryos, indicating a loss of miR-430 activity. **c**, *In situ* hybridization for maternal miR-430 target gene *cd82b*. At shield stage, *cd82b* is cleared from wild-type and MZ*pou5f1* embryos. Combined Nanog, SoxB1 and Pou5f1 LOF causes a failure in clearance (MZ*pou5f1* + Nanog + SoxB1 MO). Injection of *nanog*, *soxB1* and *pou5f1* mRNA rescues the phenotype (MO + mRNA). **d**, Cumulative plots showing the effect of each LOF condition on miR-430 target repression, as in ref. 16,

using Total RNA-seq. Plots show the distribution of log₂ fold expression level difference for each condition relative to wild type in three groups of genes defined in ref. 16: miR-430 targets with multiple 7mer or 8mer seed target sites in their 3' UTR; miR-430 targets with a single 7mer or 8mer seed in the 3' UTR; and genes lacking miR-430 seed sites in their 3' UTRs. *P* values are for two-sided Wilcoxon rank-sum tests comparing each of the two miR-430 target groups to the non-targets. MZ*dicer* expression data are from ref. 16.

Displacement of the curve to the left (–) from the grey control line indicates a larger fraction of genes are accumulated (fail to be degraded) in the indicated condition compared to wild type. Nanog has the strongest effect, although there is also an effect from the combined loss of Pou5f1 and SoxB1. **e**, Cumulative plots showing the effect of triple LOF with and without mRNA rescue on miR-430 target repression, using poly(A)⁺ selection RNA-seq. At 6 h.p.f., miR-430 targets fail to be degraded in the LOF condition compared to wild type, with expression levels of targets high in the LOF relative to wild type. Co-injection of *nanog*, *soxB1* and *pou5f1* mRNAs restores miR-430 activity, and the targets' expression levels are restored to near wild-type levels. **f**, At 8 h.p.f., miR-430 targets are still undegraded in the LOF, but are degraded to wild-type levels in the rescue. *P* values are for two-sided Wilcoxon rank-sum tests comparing each of the two miR-430 target groups to the non-targets.



Extended Data Figure 9 | Nanog, Pou5f1 and SoxB1 bind to and regulate embryonic genes.

a, Nanog chromatin immunoprecipitation sequencing binding data in zebrafish at 3.3 h.p.f. (ref. 24) was re-analysed to determine Nanog-bound regions genome wide. Pie charts show percentage of genes in each category that are associated with Nanog bound regions (± 5 kb). 74% of first-wave genes detected at sphere were associated with Nanog binding, twofold higher than subsequent-wave genes ($P = 3.7 \times 10^{-29}$, two-sided Fisher's exact test). Low expressed zygotic genes are also less associated with Nanog-bound regions. For those genes that are nonetheless affected by Nanog LOF, this suggests that they are influenced by Nanog indirectly, rather than through Nanog binding at the gene locus. The enrichment of Nanog binding on the first-wave genes versus subsequent waves supports a model where Nanog has a central role in the regulation of the activation of the first wave of zygotic transcription. **b**, ChIP-seq data for Nanog, Oct4 and Sox2 in mouse embryonic stem cells^{57,58} were used to examine the binding profiles of genes transcribed during pre-implantation mouse embryogenesis⁵⁹, as ChIP data do not exist for early mouse embryos. Three gene groups were analysed: α -amanitin-sensitive genes expressed at early 2-cell stage (minor wave ZGA), α -amanitin sensitive genes expressed at late 2-cell stage (major wave ZGA), and genes expressed during the 4–8-cell stages (mid-preimplantation). Gene

promoters (defined to be 5 kb upstream to 50 bp downstream the annotated transcription start site of a gene) are highly enriched in binding sites among the genes comprising ZGA, as compared to the genome as a whole ($P = 4.03 \times 10^{-7}$ for the minor wave, $P = 6.05 \times 10^{-18}$ major wave, two-sided Fisher's exact test). Genomic coordinates (mm8) for genes were defined by NIA/NIH U-cluster annotations for the microarray probes in ref. 59. Note that not all of the genes expressed during ZGA are necessarily expressed in ES cells; thus, the binding proportions are likely to be underestimates. Although these represent two different states of development, these results are consistent with a role for these factors in activating the earliest waves of zygotic gene expression also in mammals. **c**, Model showing maternal gene expression in red and zygotic gene expression in blue during the maternal to zygotic transition. Gene expression is depicted on the y axis and time on the x axis. During the MZT, Nanog, SoxB1 and Pou5f1 are required to activate a large fraction of zygotic genes, including miR-430, which in turn is responsible for the clearance of a significant portion of maternal mRNAs. In the loss of function of Nanog, SoxB1 and Pou5f1, there is a reduction in zygotic gene activation, causing a failure in the establishment of the zygotic developmental program, including loss of miR-430 expression and maternal mRNA clearance.

Extended Data Table 1 | Summary of Illumina sequencing data generated in this study

Sample*	Preparation	Age†	Genotype	Treatment	Total reads	rRNA	Aligned‡
1	input mRNA	2	WT	none	11,701,690	7,122,193	3,601,785
2	ribosome profiling	2	WT	none	35,324,638	28,557,085	4,782,034
3	input mRNA	2	WT	Nanog MO, SoxB1 MO	10,054,885	5,882,165	3,376,709
4	ribosome profiling	2	WT	Nanog MO, SoxB1 MO	37,708,163	28,354,953	5,946,384
5	RiboZero	2	WT	none	13,290,599	6,830,823	4,757,404
6	poly(A)+	4	WT	poly(A)+	21,504,328	NA	17,269,920
7	RiboZero	4	WT	none	49,104,024	29,072,153	16,633,109
8	RiboZero	4	WT	a-amanitin	43,280,984	17,159,279	22,541,771
9	RiboZero	4	WT	cycloheximide	60,496,090	13,980,195	40,960,186
10	RiboZero	4	WT	U1U2 MO	57,668,297	37,115,620	16,937,564
11	RiboZero	4	WT	Nanog MO	15,630,076	6,248,360	7,983,685
12	RiboZero	4	WT	SoxB1 MO	17,468,157	8,655,861	7,315,193
13	RiboZero	4	WT	Nanog MO, SoxB1 MO	13,583,155	6,597,214	5,853,123
14	RiboZero	4	MZpou5f1	none	116,396,185	90,173,314	20,274,383
15	RiboZero	4	MZpou5f1	Nanog MO	91,577,210	45,269,682	39,254,068
16	RiboZero	4	MZpou5f1	SoxB1 MO	47,420,118	32,192,741	12,137,699
17	RiboZero	4	MZpou5f1	Nanog MO, SoxB1 MO	42,220,676	28,214,894	11,452,962
18	RiboZero	4	MZpou5f1	Nanog MO, SoxB1 MO, rescue mRNA	63,785,933	22,119,249	36,078,935
19a	RiboZero	6	WT	none	14,503,666	5,448,147	7,487,251
19b	RiboZero	6	WT	none	15,074,846	8,338,535	5,303,876
19c	RiboZero	6	WT	none	17,682,683	8,153,773	7,806,485
20	poly(A)+	6	WT	none	22,626,103	NA	20,010,462
21	RiboZero	6	WT	a-amanitin	35,748,801	3,075,825	9,151,010
22	RiboZero	6	WT	cycloheximide	11,123,998	3,691,903	6,742,954
23	RiboZero	6	WT	Nanog MO	16,430,596	7,745,144	7,096,717
24a	RiboZero	6	WT	Nanog MO, SoxB1 MO	14,084,576	8,615,769	4,263,310
24b	RiboZero	6	WT	Nanog MO, SoxB1 MO	14,567,957	7,517,631	5,664,836
25	RiboZero	6	MZpou5f1	none	101,366,092	81,625,522	13,520,349
26	RiboZero	6	MZpou5f1	SoxB1 MO	13,616,658	5,839,148	6,383,511
27	RiboZero	6	MZpou5f1	Nanog MO, SoxB1 MO	28,543,110	13,273,679	12,670,402
28	poly(A)+	6	MZpou5f1	Nanog MO, SoxB1 MO	25,148,861	NA	22,263,359
29	poly(A)+	6	MZpou5f1	Nanog MO, SoxB1 MO, rescue mRNA	23,785,791	NA	21,033,046
30	poly(A)+	8	WT	none	23,504,890	NA	20,790,090
31	poly(A)+	8	MZpou5f1	Nanog MO, SoxB1 MO	25,758,851	NA	22,615,585
32	poly(A)+	8	MZpou5f1	Nanog MO, SoxB1 MO, rescue mRNA	23,475,791	NA	20,291,649

* All rows represent separately collected biological samples; that is, 19a, 19b and 19c, and 24a and 24b are biological replicates.

† Age in hours post fertilization.

‡ Reads aligning to the genome, minus rRNA-aligning reads where applicable.

Activated ClpP kills persisters and eradicates a chronic biofilm infection

B. P. Conlon¹, E. S. Nakayasu^{2†}, L. E. Fleck¹, M. D. LaFleur³, V. M. Isabella¹, K. Coleman³, S. N. Leonard⁴, R. D. Smith², J. N. Adkins² & K. Lewis¹

Chronic infections are difficult to treat with antibiotics but are caused primarily by drug-sensitive pathogens. Dormant persister cells that are tolerant to killing by antibiotics are responsible for this apparent paradox. Persisters are phenotypic variants of normal cells and pathways leading to dormancy are redundant, making it challenging to develop anti-persister compounds. Biofilms shield persisters from the immune system, suggesting that an antibiotic for treating a chronic infection should be able to eradicate the infection on its own. We reasoned that a compound capable of corrupting a target in dormant cells will kill persisters. The acyldepsipeptide antibiotic (ADEP4) has been shown to activate the ClpP protease, resulting in death of growing cells. Here we show that ADEP4-activated ClpP becomes a fairly nonspecific protease and kills persisters by degrading over 400 proteins, forcing cells to self-digest. Null mutants of *clpP* arise with high probability, but combining ADEP4 with rifampicin produced complete eradication of *Staphylococcus aureus* biofilms *in vitro* and in a mouse model of a chronic infection. Our findings indicate a general principle for killing dormant cells—activation and corruption of a target, rather than conventional inhibition. Eradication of a biofilm in an animal model by activating a protease suggests a realistic path towards developing therapies to treat chronic infections.

The current antibiotic crisis stems from two distinct phenomena, drug resistance and drug tolerance. Resistance mechanisms such as drug efflux or modification prevent antibiotics from binding to their targets¹, allowing pathogens to grow. Antibiotic tolerance is the property of persister cells, phenotypic variants of regular bacteria². Antibiotics kill by corrupting their targets, but these are inactive in dormant persisters, leading to tolerance^{3,4}. Persisters were discovered by Joseph Bigger in 1944, when he found that a small sub-population of *Staphylococcus aureus* survives treatment with penicillin⁵. We identified persisters as the main component responsible for drug tolerance of biofilms⁶. A multitude of chronic diseases is associated with biofilms: endocarditis, osteomyelitis, infections of catheters and indwelling devices, gingivitis and deep-seated infections of soft tissues^{7,8}. In *Escherichia coli*, which has served as a model organism for studying persisters, pathways leading to dormancy are highly redundant and largely depend on the action of toxin/antitoxin modules^{3,9}. Protein synthesis inhibition by the HipA toxin^{10,11}, a kinase¹² that phosphorylates glutamyl-transfer RNA synthetase GltX¹¹, and by at least 10 different messenger RNA endonucleases such as RelE, MazF and YafQ^{3,9,13,14} leads to dormancy. Damage of DNA induces the SOS response and expression of the TisB toxin¹⁵, which is an endogenous antimicrobial peptide¹⁶ and causes persister formation by opening an ion channel¹⁷. This decreases the proton motive force and ATP levels, leading to target shutdown and a dormant, drug-tolerant state. The multiplicity of dormancy pathways precludes development of drugs that could prevent persister formation¹⁸.

We reasoned that a compound capable of corrupting a target in dormant, energy-deprived cells will kill persisters. Acyldepsipeptide (ADEP) activates the ClpP protease, and it was reported to kill growing cells¹⁹.

Normally, ClpP recognizes and eliminates misfolded proteins with the aid of ATP-dependent ClpX, C or A subunits²⁰. ADEP binds to ClpP and keeps the catalytic chamber open, allowing entry to peptides and proteins^{21,22}. In the presence of ADEP, proteolysis by ClpP no longer

depends on ATP²³. Several related ADEP compounds are produced by *Streptomyces hawaiiensis*²⁴, and a more potent derivative, ADEP4 (Fig. 1), showed good activity against a variety of Gram-positive bacteria¹⁹. ADEP4 was efficacious in a lethal systemic murine infection of *Enterococcus faecalis* and *S. aureus* and in lethal sepsis caused by *Streptococcus pneumoniae* in the rat¹⁹. Nascent polypeptides emerging from the ribosome, rather than mature folded proteins, were proposed to be primary targets of ADEP4/ClpP²³. This would indicate that ADEP4 targets growing cells with active protein synthesis. A particular mature protein, FtsZ, has been reported to be a major target of ADEP4/ClpP²⁵. FtsZ forms the cell division ring, suggesting activity of ADEP4 against growing cells as well.

Here we sought to examine the ability of ADEP4 to activate protein degradation in non-growing cells and find that in its presence, ClpP becomes a fairly nonspecific protease. Null *clpP* mutants are resistant to ADEP4 (ref. 19), but we find that they are highly susceptible to killing by a variety of antibiotics. Combining ADEP4 with rifampicin leads to eradication of persisters in growing, stationary and biofilm populations of *S. aureus in vitro*, and clears a deep-seated murine biofilm infection that is untreatable with conventional antibiotics.

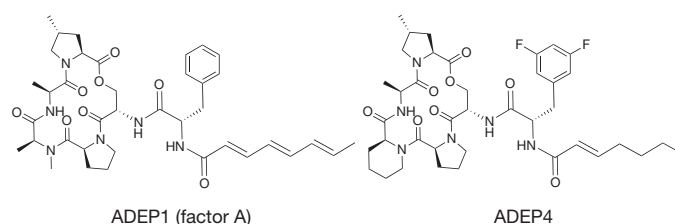


Figure 1 | Structures of acyldepsipeptide factor A and its synthetic derivative ADEP4.

¹Antimicrobial Discovery Center, Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA. ²Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, USA. ³Arietis Corporation, Boston, Massachusetts 02118, USA. ⁴Bouvé College of Health Sciences, School of Pharmacy, Northeastern University, Boston, Massachusetts 02115, USA. [†]Present address: Bindley Bioscience Center, Purdue University, West Lafayette, Indiana 47907, USA.

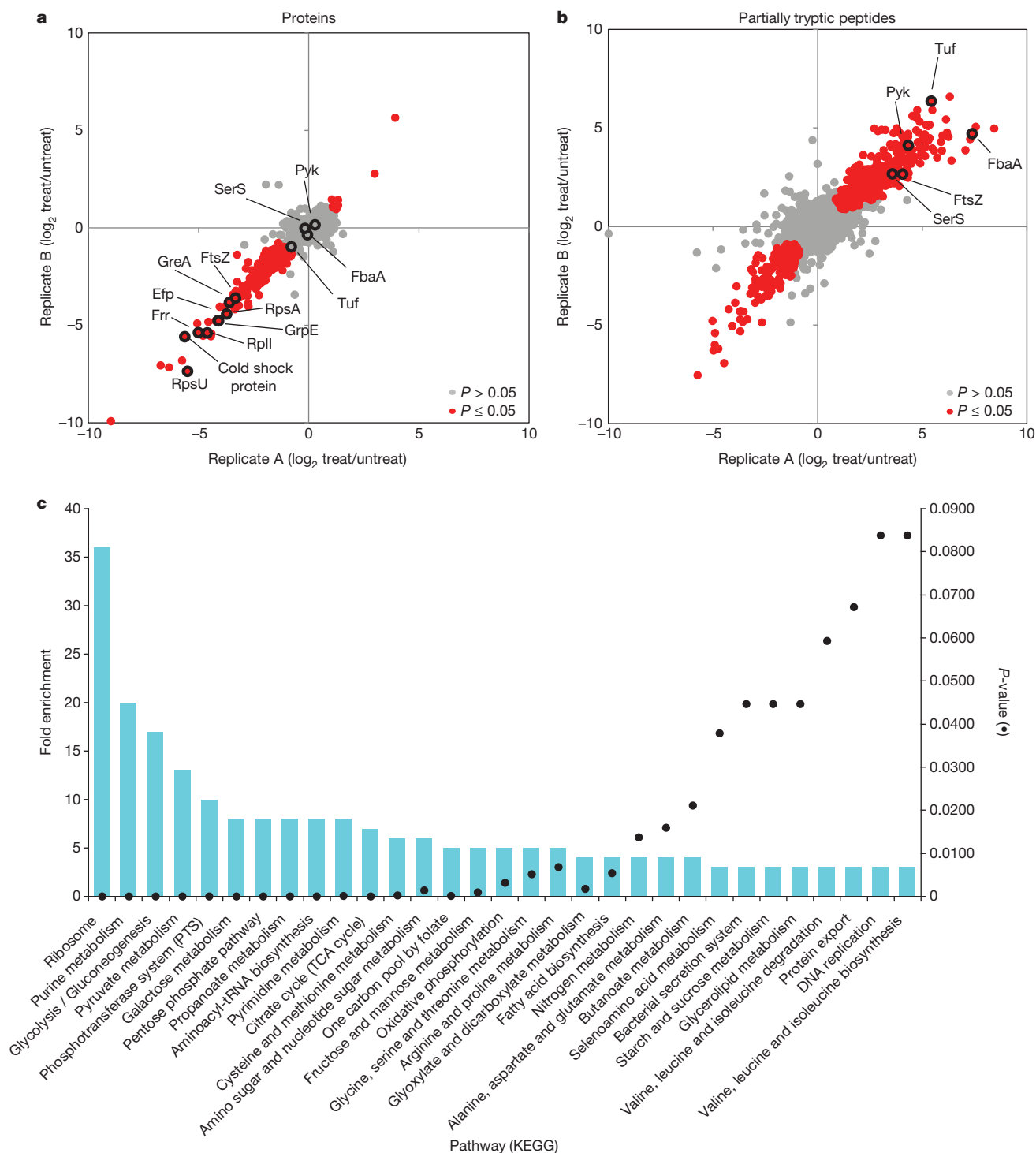


Figure 2 | Quantitative proteomic analysis of *S. aureus* cells treated with ADEP4 reveals extensive protein degradation. *S. aureus* cells were treated with ADEP4 in biological duplicates and submitted for global quantitative proteomic analysis. **a**, **b**, The dispersion graphs show the relative abundances (treated/untreated) of total proteins (**a**) and partially tryptic peptides (**b**) in different biological replicates ($n = 2$). The significant changes in abundances ($P \leq 0.05$ and > 2 -fold) are represented in red circles.

ADEP4 causes extensive protein degradation

Previous studies showing that ADEP targets nascent peptides and FtsZ in particular were performed with short exposure times and with rapidly growing cells, and we considered the possibility that longer incubation with ADEP may result in nonspecific degradation of proteins in non-growing cells. A stationary phase population of *S. aureus* was chosen to

c, Function-enrichment analysis of proteins degraded by ADEP4. Functions overrepresented among proteins degraded by ADEP4 were annotated using Database for Annotation, Visualization and Integrated Discovery (DAVID) and the overrepresented pathways compared to the genome background are shown as columns, whereas their P -values are represented by the black dots. Bayesian moderated t -test was used to provide P -values that were further corrected by the data set size.

test this, as cells are not dividing and synthesis of nascent polypeptides is strongly downregulated²⁶. Stationary cells of methicillin-resistant *S. aureus* (MRSA) were exposed to ADEP4 for 24 h and the resulting proteome was compared with that of an untreated control (Fig. 2).

Proteomic analysis of untreated stationary cells led to the detection of 1,712 proteins (65% of the predicted open reading frames). Treatment

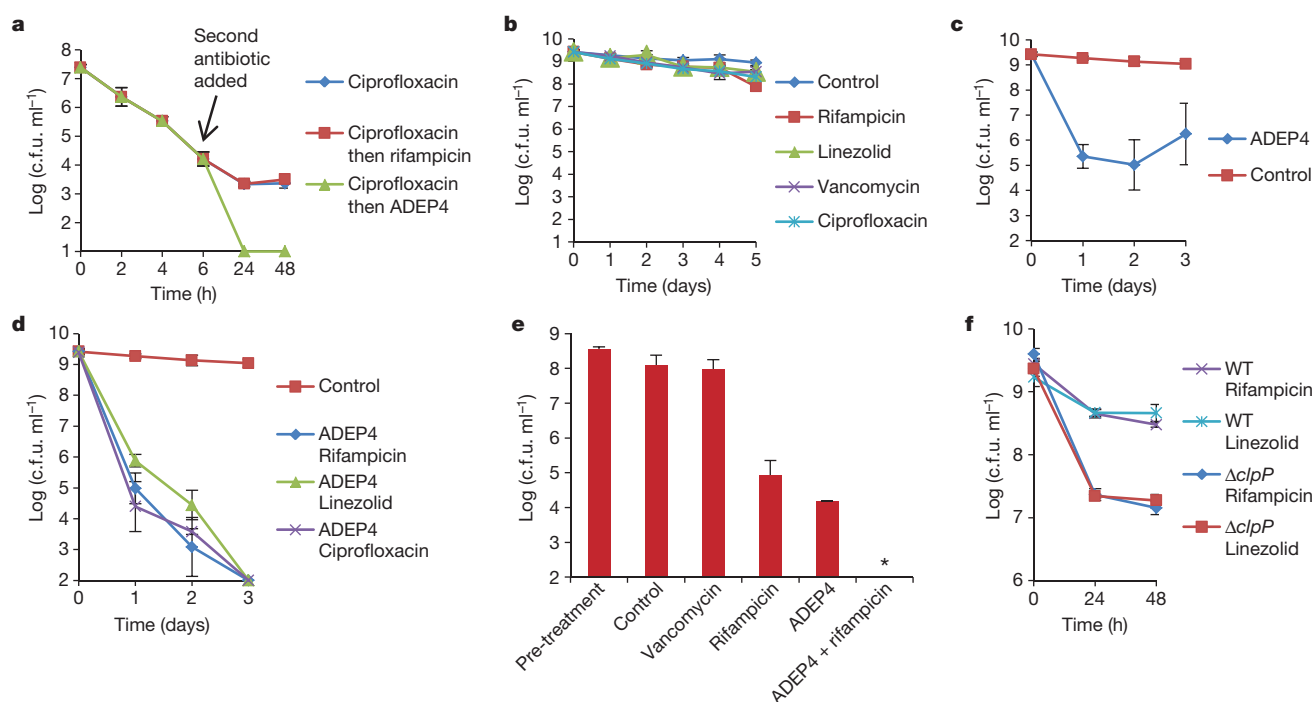


Figure 3 | ADEP4 kills persisters. **a**, ADEP4 kills persisters surviving ciprofloxacin treatment. **b**, Conventional antibiotics are inactive against stationary phase *S. aureus*. **c**, ADEP4 activity against stationary *S. aureus*. **d**, **e**, ADEP4 in combination with rifampicin, linezolid or ciprofloxacin eradicates stationary phase *S. aureus* to the detection limit in MHB

(**d**) and in 24 h in chemically defined medium (**e**). The x axis is the limit of detection. The asterisk represents eradication to the limit of detection. **f**, ADEP4 resistant mutants are less tolerant to rifampicin and linezolid than the parent wild-type strain. Data are representative of 3 independent experiments. Error bars represent s.d.

with ADEP4 resulted in decreased abundance of 243 proteins ($P \leq 0.05$ and twofold decrease) (Fig. 2a) (Supplementary Table 1). However, this is probably an underestimate. The proteome reports changes in the relative abundance of peptides produced by exogenous trypsin cleavage. A protein only cleaved once by ADEP4/ClpP, for example, would still generate several tryptically derived peptides, and not appear to show an overall decrease in protein abundance.

To address this, we examined partially tryptic peptides to uncover additional ADEP4/ClpP targets (Fig. 2b). Partially tryptic peptides exist following trypsin treatment at certain abundance in cells due to natural degradation. However, the levels of these peptides changed markedly due to degradation induced by addition of ADEP4 (red spots). An increase of partially tryptic peptides indicates ADEP4-dependent degradation of a protein. This analysis revealed 174 additional ADEP4/ClpP targets (peptides of increased abundance; Fig. 2b; Supplementary Table 2), bringing their total number to 417. A decrease on the other hand indicates that a particular degradation product, present at the time of ADEP4 addition, can be further degraded by ADEP4/ClpP, but these are of less relevance to the study.

Essential ribosomal proteins were among the most strongly diminished by ADEP4/ClpP, with proteins S21, L9, S1 and ribosomal recycling factor all showing between 17- and 64-fold reduction in the ADEP4 treated sample. Elongation factor Tu, pyruvate kinase and fructose bi-phosphate aldolase were among the proteins with the largest increase in non-trypsin cleavage sites (Fig. 2b). FtsZ was also one of the many strongly degraded proteins perhaps because of its disordered carboxy terminus²⁷. Other than the ribosome, degraded proteins belonged to various functional types, including purine metabolism, glycolysis and aminoacyl-tRNA biosynthesis, among others (Fig. 2c).

ADEP4 kills persister cells

The proteomic data indicates that ADEP4 forces the cell to self-digest, and may be effective in killing dormant cells. ADEP4 uncouples ClpP

from the requirement to use ATP, which would help kill persisters with low energy levels¹⁵. In a control experiment, ciprofloxacin was added to an exponentially growing culture of *S. aureus*, which produced a typical biphasic killing pattern with surviving persisters (Fig. 3a). Addition of rifampicin to surviving persisters had no effect on their viability, in agreement with previous observations on the multidrug tolerant nature of these cells^{3,28}. By contrast, addition of ADEP4 led to eradication of persisters to the limit of detection (Fig. 3a). Next, we examined the ability of ADEP4 to kill stationary cells of *S. aureus*. Stationary phase *S. aureus* cells behave as persisters and are extremely difficult to kill with antibiotics^{28,29}, even over a 5-day period (Fig. 3b). Furthermore, combinations of vancomycin, rifampicin and ciprofloxacin had limited activity against this population (Extended Data Fig. 1). ADEP4 showed excellent killing, decreasing the cell count of a stationary culture by 4 log₁₀ in two days (Fig. 3c), but the population rebounded after day 3. Null mutants of *clpP* are resistant to ADEP4 (ref. 19) and arise with high frequency because ClpP is not essential in *S. aureus*. No cross-resistance to marketed antibiotics was identified for ADEP4 (ref. 19). Sequencing of 9 isolates of this culture showed mutations in *clpP*, and all of them displayed the temperature-sensitive phenotype characteristic of null *clpP* mutants³⁰ (Extended Data Fig. 2). To suppress resistant mutants, ADEP4 was paired with either rifampicin, linezolid or ciprofloxacin. ADEP4 with rifampicin eradicated a stationary population of *S. aureus* to the limit of detection (Fig. 3d). This shows that ADEP4, unlike conventional antibiotics, has a remarkable ability to kill drug-tolerant persister cells. The rich Mueller-Hinton broth (MHB) in which these experiments were performed probably does not reflect conditions *in vivo* where pathogens experience nutrient limitation. We therefore tested susceptibility to killing of stationary cells in a chemically defined medium³¹. Killing in the minimal medium by ADEP4 with rifampicin was even more effective than in MHB, eradicating the population in 24 h (Fig. 3e). Complete sterilization in these experiments was unexpected—the frequency of *clpP* mutants is 10⁻⁶, and in a population of 10⁹ cells, there should have

been 10^3 survivors. To investigate this, a *clpP* mutant was examined for its susceptibility to linezolid and rifampicin (Fig. 3f). The $\Delta clpP$ strain had the same minimum inhibitory concentration (MIC) as the wild type, but stationary phase counts were reduced 10- to 100-fold more than the wild type by linezolid or rifampicin in stationary state. A mutation in *clpP* apparently diminishes the fitness of cells and makes them vulnerable to certain antibiotics. In agreement with this, a *clpP* mutant was reported to be avirulent in a murine skin abscess model of infection³⁰. We then tested the eradicating potential of the ADEP4 and rifampicin combination against a variety of *S. aureus* strains. These included the laboratory strain SA113, as well as clinical isolates USA300, UAMS-1 and strain 37. USA300 is a community acquired MRSA and is the most common cause of staphylococcal skin and soft tissue infections in the United States³². UAMS-1 is a highly virulent clinical isolate associated with chronic osteomyelitis³³. Strain 37 was isolated from a patient undergoing vancomycin therapy who succumbed to infection³⁴. No colonies were detected in any of these strains after 72 h of incubating stationary cultures with ADEP4 and rifampicin (Fig. 4).

ADEP4 with rifampicin eradicates biofilm

Biofilms produced by the osteomyelitis-associated strain UAMS-1 displayed a similar tolerance to antibiotics as stationary phase cultures (Fig. 5). ADEP4 showed considerable killing following 24 h of treatment, but the population rebounded after 72 h. Again, a combination of ADEP4 with rifampicin resulted in eradication of living cells in the biofilm to the limit of detection (Fig. 5). The replacement of antibiotics with fresh medium did not result in re-growth after 3 days of ADEP4 and rifampicin treatment, confirming the complete eradication of living cells. An elimination of a biofilm is unprecedented for such low, clinically achievable concentrations of compounds.

ADEP4 with rifampicin eradicates infection

Eradication of stationary and biofilm populations was an encouraging sign that ADEP4 could be a very useful antibiotic against untreatable chronic infections. To test this, we used a deep-seated mouse thigh infection model. In a standard thigh model, a mouse is infected with a low dose of pathogen and antibiotic therapy begins within a few hours of infection. Under these conditions, conventional antibiotics are very effective. In the deep-seated model, the mouse is made neutropenic by treatment with cyclophosphamide, a large dose of pathogen is delivered and the infection is allowed to develop for 24 h before therapy, leading to a severe, recalcitrant, deep-seated infection. This model emulates a difficult to treat human deep-seated chronic infection in

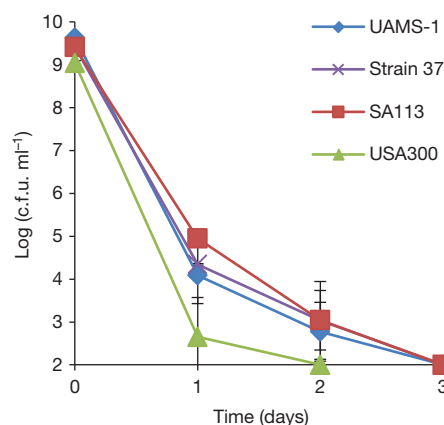


Figure 4 | ADEP4 with rifampicin eradicates a variety of *S. aureus* strains. *S. aureus* was grown in MHB for 16 h and challenged with $10\times$ MIC of ADEP4 and rifampicin. Colony counts were performed every 24 h. The x axis is the limit of detection. Data are representative of 3 independent experiments. Error bars represent s.d.

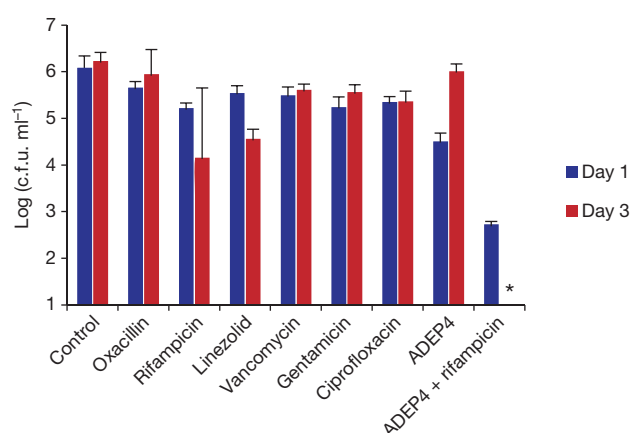


Figure 5 | ADEP4 kills a *S. aureus* biofilm and in combination with rifampicin eradicates the population. The x axis is the limit of detection. An asterisk represents eradication to the limit of detection. Data are representative of 3 independent experiments. Error bars represent s.d.

immunocompromised patients. We performed histopathology of the infected thigh and detected massive aggregates of *S. aureus* cells with Gram staining (Fig. 6a). Electron microscopy of cross-sections of the infected tissue revealed *S. aureus* growing in biofilms adhered to muscle cells (Fig. 6a). Administration of vancomycin, rifampicin or a combination of both decreased the viable counts, but did not clear the infection (Fig. 6b). Furthermore, no notable difference was observed between mice treated for 24 h or 48 h with vancomycin in this model, indicating the presence of a persister subpopulation surviving the antibiotic treatment (Fig. 6b). Remarkably, an ADEP4 and rifampicin combination led to sterilization of the infected tissue to the limit of detection within 24 h (Fig. 6c). Based on this efficacious dose and the mouse pharmacokinetics data¹⁹, we performed a hollow-fibre experiment and found that the combination of ADEP4 and rifampicin also resulted in complete eradication of the pathogen to the limit of detection (Extended Data Fig. 3).

Discussion

The rise in biofilm infections is a recent phenomenon, mainly a side-effect of medical intervention³⁵. Biofilms form readily on indwelling devices such as catheters, prostheses and heart valves. Biofilms have a complex architecture and developmental program^{36,37} and form a protective environment for persisters, shielding them from the immune system. In patients undergoing cancer chemotherapy or organ transplantation or in the elderly, the immune system is compromised, enabling deep-seated infections in soft tissues to take hold. Even disseminating infections of *S. aureus* are difficult to eradicate in immunocompromised patients. The dormant state of persisters and the multiplicity of the pathways leading to their formation make treatment of chronic infections unusually challenging. Our results demonstrate that persister cells in a biofilm can be killed with a protease-activating antibiotic. This study shows that persisters are not invulnerable, and helps settle an important uncertainty surrounding chronic diseases—it has been unclear whether conventional antibiotics fail owing to their ineffective killing or simply because they do not reach all pathogens at the site of infection. We had previously described high-persister (*hip*) mutants that are selected in the course of antibiotic treatment in patients with *Candida albicans* biofilms³⁸ or with *Pseudomonas aeruginosa* in the lungs of patients with cystic fibrosis³⁹. Selection for increased production of persister cells suggests that antibiotics effectively reach the pathogens. Sterilization of a deep-seated biofilm infection with ADEP4, but not with conventional antibiotics, shows directly that the problem indeed lies in pathogen tolerance. Pathogens surviving antibiotic treatment in a chronic infection are detrimental not only to a given patient, but to society as well. A large, lingering population of pathogens is

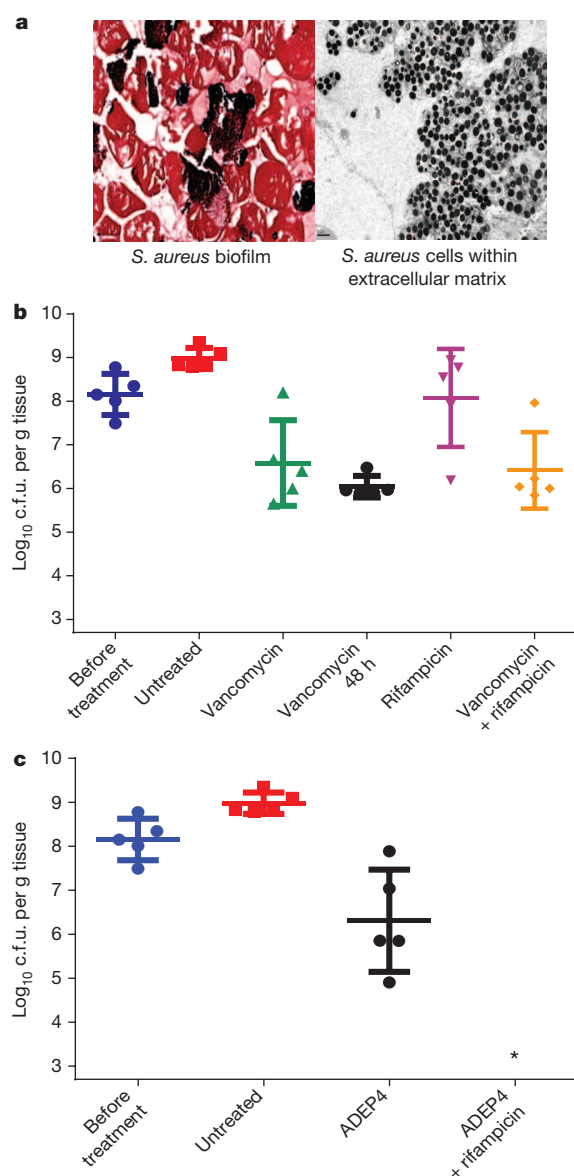


Figure 6 | ADEP4 in combination with rifampicin eradicates a deep-seated mouse biofilm infection. **a**, Histopathology of *S. aureus* infected thighs reveals the presence of a biofilm. Gram staining $\times 80$ magnification (left); electron micrograph $\times 8,000$ magnification (right). **b**, Single day (rifampicin 30 mg per kg once, vancomycin 110 mg per kg twice) treatments with rifampicin and vancomycin. A second day of vancomycin treatment (vancomycin 48 h) reveals an antibiotic tolerant subpopulation. **c**, Single day ADEP4 rifampicin combination eradicates *S. aureus* in the deep-seated infection. An asterisk represents eradication to the limit of detection. Groups of 5 neutropenic Swiss mice were used for each experiment. Colony-forming units (c.f.u.) from each mouse are plotted as individual points and error bars represent the deviation in c.f.u. within an experimental group.

fertile ground for the development of resistance^{40–42}. The ability to efficiently eradicate an infection will help reduce the spread of resistance.

ADEP4 is remarkable in its ability to kill dormant cells. Persisters formed by competitors would present an obvious problem for Actinomycetes, and it is perhaps not surprising that they evolved compounds capable of killing both growing and dormant cells. ADEP4 points to a general principle of killing, activation and corruption of a target (Extended Data Fig. 4). Apart from ADEP4, other activators of ClpP⁴³ may be developed into therapeutics, and additional bacterial proteases such as Lon could be used as targets for killing specialized survivor cells. This general principle of killing may be applied to other organisms as well

and prove effective in developing therapeutics to treat fungal infections and cancer.

METHODS SUMMARY

In all experiments, bacterial cells were cultured in 20 ml of Mueller-Hinton (MH), brain heart infusion (BHI) broth or chemically defined media³¹ at 37 °C and were aerated at 225 r.p.m in 250 ml flasks. Antibiotics were applied at the following concentrations, corresponding to 10 \times MIC: vancomycin 10 $\mu\text{g ml}^{-1}$, ADEP4 5 $\mu\text{g ml}^{-1}$, rifampicin 0.4 $\mu\text{g ml}^{-1}$, linezolid 10 $\mu\text{g ml}^{-1}$ and ciprofloxacin 3 $\mu\text{g ml}^{-1}$. MICs were the same for each strain examined. The strains used in this study were *S. aureus*: ATCC 33591, UAMS-1, USA300, SA113 and strain 37 (ref. 34). Biofilm survival assays were performed in 96-well polystyrene plates in BHI broth in a static 37 °C incubator. Biofilm was allowed to develop for 24 h. Wells were gently washed with PBS and fresh medium containing antibiotics was added to each well. Biofilms were incubated for either 24 or 72 h. Biofilms were then washed and sonicated in PBS. Serial dilutions were performed and 10 μl aliquots were spotted on BHI agar. iTRAQ proteomics was performed on stationary phase ATCC 33591 treated with 10 \times MIC of ADEP4 for 24 h and compared to an untreated control. Mouse experiments were performed with female 6-week-old Swiss-Webster (Taconic) mice that were first rendered neutropenic by cyclophosphamide treatment (150 mg per kg and 100 mg per kg 96 h and 24 h before infection, respectively). Infection was allowed to develop for 24 h before commencement of antibiotic therapy. All antibiotics were delivered intraperitoneally. Animal experiments were carried out at Northeastern University and conformed to institutional animal care and use policies.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 August; accepted 18 October 2013.

Published online 13 November 2013.

- Alekshun, M. N. & Levy, S. B. Molecular mechanisms of antibacterial multidrug resistance. *Cell* **128**, 1037–1050 (2007).
- Lewis, K. Persister cells. *Annu. Rev. Microbiol.* **64**, 357–372 (2010).
- Keren, I., Shah, D., Spoering, A., Kaldalu, N. & Lewis, K. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. *J. Bacteriol.* **186**, 8172–8180 (2004).
- Keren, I., Wu, Y., Inocencio, J., Mulcahy, L. R. & Lewis, K. Killing by bactericidal antibiotics does not depend on reactive oxygen species. *Science* **339**, 1213–1216 (2013).
- Bigger, J. W. Treatment of staphylococcal infections with penicillin. *Lancet* **244**, 497–500 (1944).
- Spoering, A. L. & Lewis, K. Biofilms and planktonic cells of *Pseudomonas aeruginosa* have similar resistance to killing by antimicrobials. *J. Bacteriol.* **183**, 6746–6751 (2001).
- Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: A common cause of persistent infections. *Science* **284**, 1318–1322 (1999).
- Kostakioti, M., Hadjifrangiskou, M. & Hultgren, S. J. Bacterial biofilms: development, dispersal, and therapeutic strategies in the dawn of the postantibiotic era. *Cold Spring Harb. Perspect. Med.* **3**, a010306 (2013).
- Maisonneuve, E., Shakespeare, L. J., Jorgensen, M. G. & Gerdes, K. Bacterial persistence by RNA endonucleases. *Proc. Natl Acad. Sci. USA* **108**, 13206–13211 (2011).
- Moyed, H. S. & Bertrand, K. P. *hipA*, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* **155**, 768–775 (1983).
- Germain, E., Castro-Roa, D., Zenkin, N. & Gerdes, K. Molecular mechanism of bacterial persistence by *HipA*. *Mol. Cell* **52**, 248–254 (2013).
- Correia, F. F. et al. Kinase activity of overexpressed *HipA* is required for growth arrest and multidrug tolerance in *Escherichia coli*. *J. Bacteriol.* **188**, 8360–8367 (2006).
- Harrison, J. J. et al. The chromosomal toxin gene *yafQ* is a determinant of multidrug tolerance for *Escherichia coli* growing in a biofilm. *Antimicrob. Agents Chemother.* **53**, 2253–2258 (2009).
- Maisonneuve, E., Castro-Camargo, M. & Gerdes, K. (p)ppGpp controls bacterial persistence by stochastic induction of toxin-antitoxin activity. *Cell* **154**, 1140–1150 (2013).
- Dörr, T., Vulić, M. & Lewis, K. Ciprofloxacin causes persister formation by inducing the TisB toxin in *Escherichia coli*. *PLoS Biol.* **8**, e1000317 (2010).
- Unoson, C. & Wagner, E. A small SOS-induced toxin is targeted against the inner membrane in *Escherichia coli*. *Mol. Microbiol.* **70**, 258–270 (2008).
- Guiney, P. A., Ortenberg, R., Dörr, T., Lewis, K. & Bezrukov, S. M. Persister-promoting bacterial toxin TisB produces anion-selective pores in planar lipid bilayers. *FEBS Lett.* **586**, 2529–2534 (2012).
- Lewis, K. Platforms for antibiotic discovery. *Nature Rev. Drug Discov.* **12**, 371–387 (2013).
- Brötz-Oesterhelt, H. et al. Dysregulation of bacterial proteolytic machinery by a new class of antibiotics. *Nature Med.* **11**, 1082–1087 (2005).

20. Gottesman, S., Roche, E., Zhou, Y. & Sauer, R. T. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338–1347 (1998).
21. Li, D. H. *et al.* Acyldepsipeptide antibiotics induce the formation of a structured axial channel in ClpP: A model for the ClpX/ClpA-bound state of ClpP. *Chem. Biol.* **17**, 959–969 (2010).
22. Lee, B. G. *et al.* Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nature Struct. Mol. Biol.* **17**, 471–478 (2010).
23. Kirstein, J. *et al.* The antibiotic ADEP reprogrammes ClpP, switching it from a regulated to an uncontrolled protease. *EMBO Mol. Med.* **1**, 37–49 (2009).
24. Michel, K. H. & Kastner, R. E. A54556 antibiotics and process for production thereof. US patent 4,492 650 (1985).
25. Sass, P. *et al.* Antibiotic acyldepsipeptides activate ClpP peptidase to degrade the cell division protein FtsZ. *Proc. Natl Acad. Sci. USA* **108**, 17474–17479 (2011).
26. Michalik, S. *et al.* Proteolysis during long-term glucose starvation in *Staphylococcus aureus* COL. *Proteomics* **9**, 4468–4477 (2009).
27. Buske, P. J. & Levin, P. A. A flexible C-terminal linker is required for proper FtsZ assembly *in vitro* and cytokinetic ring formation *in vivo*. *Mol. Microbiol.* **89**, 249–263 (2013).
28. Keren, I., Kaldalu, N., Spoering, A., Wang, Y. & Lewis, K. Persister cells and tolerance to antimicrobials. *FEMS Microbiol. Lett.* **230**, 13–18 (2004).
29. Johnson, P. J. & Levin, B. R. Pharmacodynamics, population dynamics, and the evolution of persistence in *Staphylococcus aureus*. *PLoS Genet.* **9**, e1003123 (2013).
30. Frees, D., Qazi, S. N., Hill, P. J. & Ingmer, H. Alternative roles of ClpX and ClpP in *Staphylococcus aureus* stress tolerance and virulence. *Mol. Microbiol.* **48**, 1565–1578 (2003).
31. Hussain, M., Hastings, J. G. & White, P. J. A chemically defined medium for slime production by coagulase-negative staphylococci. *J. Med. Microbiol.* **34**, 143–147 (1991).
32. Strykowski, M. E. & Chambers, H. F. Skin and soft-tissue infections caused by community-acquired methicillin-resistant *Staphylococcus aureus*. *Clin. Infect. Dis.* **46** (Suppl 5), S368–S377 (2008).
33. Gillaspay, A. F. *et al.* Role of the accessory gene regulator (*agr*) in pathogenesis of staphylococcal osteomyelitis. *Infect. Immun.* **63**, 3373–3380 (1995).
34. Miyazaki, M. *et al.* Vancomycin bactericidal activity as a predictor of 30-day mortality in patients with methicillin-resistant *Staphylococcus aureus* bacteremia. *Antimicrob. Agents Chemother.* **55**, 1819–1820 (2011).
35. Lewis, K. Antibiotics: recover the lost art of drug discovery. *Nature* **485**, 439–440 (2012).
36. Hung, C. *et al.* *Escherichia coli* biofilms have an organized and complex extracellular matrix structure. *mBio* **4**, e00645–13 (2013).
37. Vlamakis, H., Chai, Y., Beauregard, P., Losick, R. & Kolter, R. Sticking together: building a biofilm the *Bacillus subtilis* way. *Nature Rev. Microbiol.* **11**, 157–168 (2013).
38. LaFleur, M. D., Qi, Q. & Lewis, K. Patients with long-term oral carriage harbor high-persister mutants of *Candida albicans*. *Antimicrob. Agents Chemother.* **54**, 39–44 (2010).
39. Mulcahy, L. R., Burns, J. L., Lory, S. & Lewis, K. Emergence of *Pseudomonas aeruginosa* strains producing high levels of persister cells in patients with cystic fibrosis. *J. Bacteriol.* **192**, 6191–6199 (2010).
40. Mwangi, M. M. *et al.* Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **104**, 9451–9456 (2007).
41. Levin, B. R. & Rozen, D. E. Non-inherited antibiotic resistance. *Nature Rev. Microbiol.* **4**, 556–562 (2006).
42. Howden, B. P. *et al.* Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Pathog.* **7**, e1002359 (2011).
43. Leung, E. *et al.* Activators of cylindrical proteases as antimicrobials: identification and development of small molecule activators of ClpP protease. *Chem. Biol.* **18**, 1167–1178 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Wright and C. Blinn of AstraZeneca for assisting with the establishment of the mouse infection model, R. E. Lee, M. Pollastri and Z. Maglika for critical discussions and advice, I. Keren and S. Rowe for reading of the manuscript, H. Brewer, V. Petyuk and D. Camp II for assistance with proteomics, and Z. Zheng for assistance with ChemDraw. This work was supported by NIH award T-RO1 AI085585 to K.L., by Arietis Corporation to M.D.L. and K.C., by the NIH-NIAID IAA Y1-AI-8401 to J.N.A. and P41 GM103493-11 to R.D.S. Proteomic analysis was performed in the EMSL, a DOE-BER national scientific user facility at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under contract DE-AC05-76RLO 1830.

Author Contributions B.P.C., M.D.L., K.C. and K.L. designed the study, analysed results and wrote the manuscript. B.P.C. performed *in vitro* antibiotic susceptibility assays, collected samples for proteomics, and performed biofilm susceptibility studies and mouse infection models. V.M.I. assisted with *in vitro* susceptibility assays. E.S.N. and J.N.A. performed i-TRAQ proteomics and analysed results. L.E.F. participated in mouse infection model experiments. S.N.L. performed hollow-fibre experiments. M.D.L. was responsible for histopathology. R.D.S. provided the proteomics measurement capabilities.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.L. (k.lewis@neu.edu).

METHODS

Bacterial strains, plasmids, media and growth conditions. Methicillin-resistant *Staphylococcus aureus* strain ATCC 33591 was used for proteome analysis, antibiotic killing assays and *in vivo* infections. USA300, SA113, UAMS-1 and strain 37 (ref. 34) were also used in antibiotic killing assays. Biofilm experiments were carried out with UAMS-1. Stationary phase populations were prepared as follows: bacteria from frozen stock were grown in 20 ml of Mueller-Hinton broth (MHB) or in a chemically defined medium³¹ in 250 ml conical flasks with aeration at 225 r.p.m. at 37 °C overnight. Exponential phase cultures were prepared as follows: a stationary overnight culture was diluted 1:1,000 in MHB and incubated at 37 °C with aeration at 225 r.p.m. until $A_{600\text{ nm}} = 0.5$ was reached. Biofilms were grown in brain heart infusion (BHI) broth. Mueller-Hinton agar (MHA) and BHI agar were used for colony counts.

Antibiotic susceptibility assays. Overnight, stationary phase, biological triplicates were used in all susceptibility assays. Bacteria were incubated in the presence of antibiotics at 37 °C with aeration at 225 r.p.m. Antibiotic concentrations, corresponding to 10× the minimum inhibitory concentration were as follows: vancomycin 10 µg ml⁻¹, ADEP4 5 µg ml⁻¹, rifampicin 0.4 µg ml⁻¹, linezolid 10 µg ml⁻¹ and ciprofloxacin 3 µg ml⁻¹. Live cell numbers at a given time point were determined as follows: 100 µl of culture was removed from the flask and centrifuged at 10,000g for 1 min. The resulting pellet was resuspended in PBS. Serial dilutions were performed and 10 µl of each dilution was spotted onto Mueller-Hinton agar plates. Plates were allowed to dry and then incubated overnight at 37 °C.

Biofilm assays. Overnight, stationary phase, biological triplicates of UAMS-1 were diluted 1:20 in BHI broth. Then 100 µl of this culture was added to each well of a tissue-culture treated polystyrene 96-well plate. Plates were incubated at 37 °C for 24 h. Medium was carefully removed and wells were gently washed twice with PBS. Then 100 µl of fresh medium containing 10× MIC of antibiotics was carefully added to each well. Plates were incubated for either 24 or 72 h. Medium was carefully removed and wells were gently washed twice with PBS. Then 100 µl of PBS was added to each well and biofilm was solubilized by sonication for 5 minutes in a sonicating water bath (Fischer Scientific FS30). Serial dilutions of each well were performed and 10 µl of each dilution was spotted onto BHI plates and incubated overnight at 37 °C.

Proteomic analysis. Stationary phase cultures of MRSA cells were treated with 10× MIC of ADEP4 for 24 h at 37 °C. Biological duplicates of untreated control and ADEP4-treated cells were harvested and lysed in 100 mM NH₄HCO₃, 1 mM PMSF, 2 mM *N*-ethylmaleimide (NEM) and 5 mM EDTA, by mechanical disruption by vigorous vortexing in the presence of 0.1 mm diameter silica/zirconia beads. A buffer exchange was performed on the cell lysates through an Amicon 10-kDa MWCO filter into 100 mM NH₄HCO₃. The lysate was denatured/reduced in 100 mM NH₄HCO₃, 8 M urea, 5 mM DTT for 30 min at 60 °C, and then diluted to obtain a final concentration of 1 M urea, and digested with trypsin for 3 h at 37 °C. The resulting peptides were desalted using C18 SPE cartridges (Discovery C18, 1 ml, 50 mg, Sulpeco), labelled with 4-plex iTRAQ reagent (Applied Biosystems) following the manufacturer recommendations, and each of the labelled samples was mixed in equal amounts based on total peptide concentrations measured by BCA assay (Thermo Scientific). The peptide mix was then fractionated into 96 fractions by high pH reverse phase chromatography and concatenated into 24 fractions as previously described⁴⁴, and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis in a LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Peptides were loaded into capillary LC columns (75 µm × 65 cm, Polymicro) packed with C18 beads (3 µm particles, Phenomenex) connected to a custom-made 4-column LC system⁴⁵. The elution was performed in an exponential gradient from 0–100% B solvent (solvent A: 0.1% formic acid; solvent B: 90% acetonitrile/0.1% formic acid) with a constant pressure of 10,000 psi and flow rate of ~300 nl min⁻¹. Full-MS scans were obtained for *m/z* 400–2,000 with the six most intense ions selected for HCD fragmentation using a 2 *m/z* isolation width and 45% normalized collision energy.

Raw mass spectrometry data were converted to peak lists (DTA files) using the DeconMSn⁴⁶ (version 2.2.2.2, <http://omics.pnl.gov/software/DeconMSn.php>) and searched with MSGF+⁴⁷ against the *S. aureus* COL NC 002951 (2,615 sequences), bovine trypsin and human keratin sequences (all in correct and reverse orientations, 5,362 total sequences). Searching parameters included tryptic digestion in at least one of the peptide termini (partially tryptic), 10 p.p.m. peptide mass tolerance, methionine oxidation as variable modification, and cysteine alkylation with NEM and *N* terminus and lysine labelling with iTRAQ reagent as fixed modifications. Peptides were filtered with an MSGF score ≤ 10⁻⁹, resulting ≤ 1% false-discovery rate at protein level. For the quantitative analysis, the iTRAQ report ion intensities were extracted with MASIC⁴⁸ (MS/MS Automated Selected Ion Chromatogram Generator, version v2.5.3923, <http://omics.pnl.gov/software/MASIC.php>). Peptides

yielding multiple spectra had their iTRAQ reporter ions intensities summed to remove redundancy and to improve signal to noise ratio. For protein quantification, the reporter intensities of different fully tryptic peptides belonging to the same proteins were also summed. Peptides and proteins with missing data were excluded from the analysis. Because two replicates were analysed, a Bayesian moderated *t*-test (available through 'limma' BioConductor package) was applied to determine the differentially abundant proteins.

Mouse thigh infection. Six-week-old female Swiss-Webster mice (Taconic) were used in groups of five in these experiments. A group size of five mice was chosen to provide statistically significant results based on the projected outcome of experiments. Neither randomization nor blinding was deemed necessary for this animal infection model. Mice were rendered neutropenic by cyclophosphamide therapy⁴⁹. A stationary culture of *S. aureus* ATCC 33591 was centrifuged and resuspended in PBS. Then 100 µl of a 1:100 dilution (2 × 10⁶ cells) was injected to the right thigh of each mouse. Infection was allowed to progress for 24 h and mice displayed measurable increase in thigh diameter. Mice were then treated with vancomycin (Hospira), rifampicin (Pfizer), or ADEP4 (custom synthesized by WuXi AppTec). ADEP4 and rifampicin were solubilized in 100% PEG400. Vancomycin was solubilized in water. Vancomycin was dosed intraperitoneally at 110 mg per kg every 12 h. Rifampicin was dosed intraperitoneally at 30 mg per kg every 24 h. ADEP4 was dosed intraperitoneally at 25 mg per kg followed by a second 35 mg per kg dose 4 h later. Control mice were sacrificed 24 h after infection (before treatment) and 48 h after infection (untreated). Thighs were aseptically removed and homogenized in PBS using a Bullet Blender homogenizer. Homogenates were serially diluted and samples were plated on BHI agar and incubated at 37 °C overnight. Animal experiments were carried out at Northeastern University and conformed to institutional animal care and use policies.

Microscopy. Histopathology was performed at the Boston University School of Medicine Experimental Pathology Laboratory Service Core.

For the Gram stain, infected thigh tissues were aseptically dissected and fixed overnight at 4 °C in 10% formalin. Samples were dehydrated using a graded alcohol series from 70–100%, cleared with xylene to remove the dehydrant, and infiltrated with paraffin. Processed tissue was embedded in paraffin, cut in 5-µm sections, and placed on microscope slides. Slides were baked at 67 °C for 36 min. After cooling, slides were washed twice with xylene for 5 min, twice with 100% alcohol for 5 min, twice with 95% alcohol for 2 min each, with 70% alcohol for 2 min, and left in distilled water until staining. Slides were stained using a Gram stain kit from Poly Scientific R&D Corp. Slides were stained with crystal violet for 1 min and washed thoroughly with distilled water. Next, Gram's iodine was applied for 30 s and the slides were washed thoroughly with distilled water. Slides were decolorized with Gram's decolorizer until the crystal violet was washed away. Slides were rinsed with distilled water and counterstained with Gram's safranin O counterstain. Slides were washed with distilled water and air dried before a coverslip was applied. Slides were digitized at ×40 using Ventana iScan Core AU slide scanner and viewed using Image Viewer v.3.1.

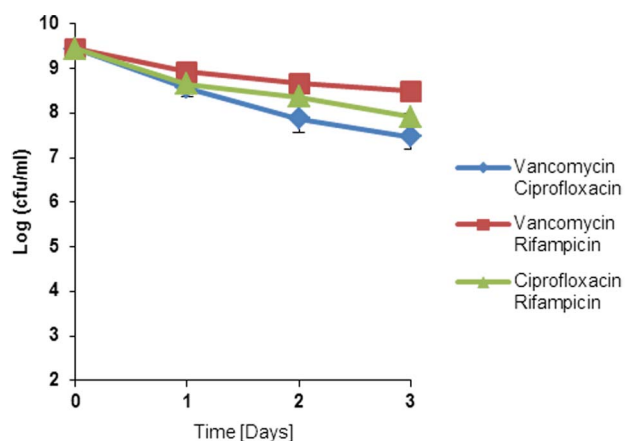
For electron microscopy, 2 mm cross-sections of infected thigh were fixed overnight at 4 °C in 2.5% glutaraldehyde/2.0% paraformaldehyde in 0.1 M sodium cacodylate buffer. Samples were post-fixed 1 h in 1.0% osmium tetroxide in 0.15 M cacodylate buffer at room temperature, dehydrated through a graded acetone series, and embedded in epoxy resin. Sections were cut at 70 nm, stained with uranyl acetate and lead citrate, and examined in a JEOL electron microscope at 80 kV. Images were recorded using a Gatan side mounted 11 megapixel digital camera.

***In vitro* hollow-fibre model.** *In vitro* pharmacokinetic/pharmacodynamic modelling experiments were performed over a 96 h period using a hollow-fibre model (Fibrecell Systems) with a culture of ~10⁷ c.f.u. ml⁻¹ as a starting inoculum. Fresh MHB was continuously supplied by a peristaltic pump (Masterflex; Cole-Parmer) set to simulate the half-lives of the antibiotics. After inoculation of the bacteria into the extracapillary space of the hollow-fibre cartridge, antibiotic was infused into the reservoir chamber via a dosing port. Free drug concentrations of vancomycin (1 g every 12 h: fC_{max} : 30 µg ml⁻¹, fC_{min} : 7.5 µg ml⁻¹, half-life: 6 h; at 50% protein binding for vancomycin these levels correspond to a total C_{max} of 60 µg ml⁻¹ and C_{min} of 15 µg ml⁻¹) and rifampicin (300 mg every 8 h: fC_{max} : 0.8 µg ml⁻¹, half-life: 3 h; at 80% protein binding for rifampicin these levels correspond to a total C_{max} of 4 µg ml⁻¹) were dosed to simulate human pharmacokinetics while ADEP4 (25 mg per kg followed by 35 mg per kg 4 h later—repeated every 24 h: C_{max} for 25 mg per kg: 11.7 µg ml⁻¹, C_{max} for 35 mg per kg: 16.4 µg ml⁻¹, half-life 1.5 h) was dosed to simulate mouse pharmacokinetics. Mouse pharmacokinetics were used for ADEP4 because there are no human pharmacokinetic data, nor are there sufficient animal pharmacokinetic data for an allometric conversion. Antibiotic regimens tested included ADEP4 alone, vancomycin alone, rifampicin alone, ADEP4 combined with rifampicin, and vancomycin combined with rifampicin. Model simulations involving two drugs with different half-lives were performed using a previously

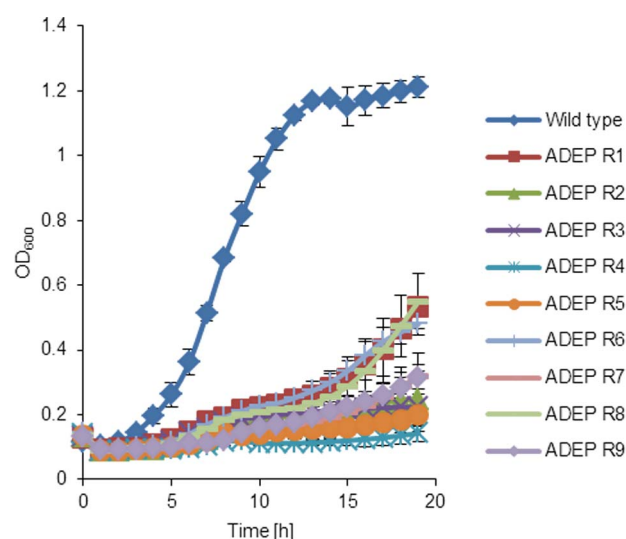
validated method⁵⁰. All experiments were performed at 37 °C in triplicate, using biological replicates, to ensure reproducibility.

Samples (1 ml) were removed at 0, 1, 2, 4, 8, 24, 28, 32, 48, 56, 72 and 96 h, serially diluted, plated on BHI agar, and incubated at 37 °C with a lower limit of detection of 10^2 c.f.u. ml⁻¹. Antibiotic concentrations were verified by bioassay using antibiotic medium 19 and *S. aureus* ATCC 33591 for ADEP4, antibiotic medium 5 and *B. subtilis* for vancomycin, and Mueller Hinton Agar and *K. rhizophila* ATCC 9341 for rifampicin. Only models using a single agent had pharmacokinetics verified while combination models were performed using the verified method described above. Pharmacokinetic parameters were analysed using WinNonlin modelling software (Pharsight). Pharmacokinetic values from the models were all within 10% of targets. Overall activity of regimens over the 96 h period was compared by calculating the area under the bacterial kill curve for each regimen using SigmaPlot software (version 11.1, Systat Software). The areas under the curve were then compared using analysis of variance (ANOVA) with Tukey's post-hoc test with IBM SPSS Statistics (Version 19.0, SPSS Inc).

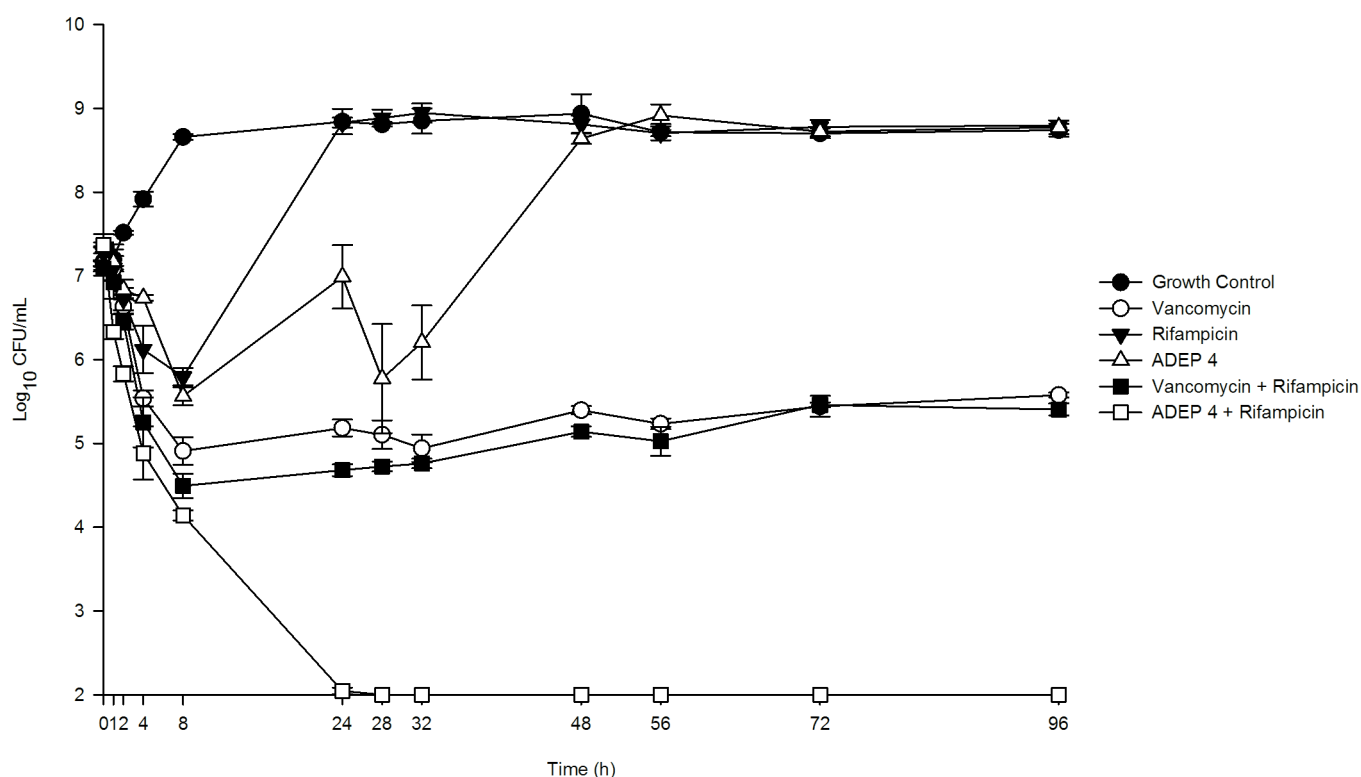
44. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019–2026 (2011).
45. Livesay, E. A. *et al.* Fully automated four-column capillary LC–MS system for maximizing throughput in proteomic analyses. *Anal. Chem.* **80**, 294–302 (2008).
46. Mayampurath, A. M. *et al.* DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **24**, 1021–1023 (2008).
47. Kim, S. *et al.* The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. *Mol. Cell Proteomics* **9**, 2840–2852 (2010).
48. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput. Biol. Chem.* **32**, 215–217 (2008).
49. Zuluaga, A. F. *et al.* Neutropenia induced in outbred mice by a simplified low-dose cyclophosphamide regimen: characterization and applicability to diverse experimental models of infectious diseases. *BMC Infect. Dis.* **6**, 55 (2006).
50. Blaser, J. *In-vitro* model for simultaneous simulation of the serum kinetics of two drugs with different half-lives. *J. Antimicrob. Chemother.* **15** (Suppl A), 125–130 (1985).



Extended Data Figure 1 | Combinations of conventional antibiotics against stationary phase *S. aureus*. Data are representative of 3 independent experiments. Error bars represent s.d.

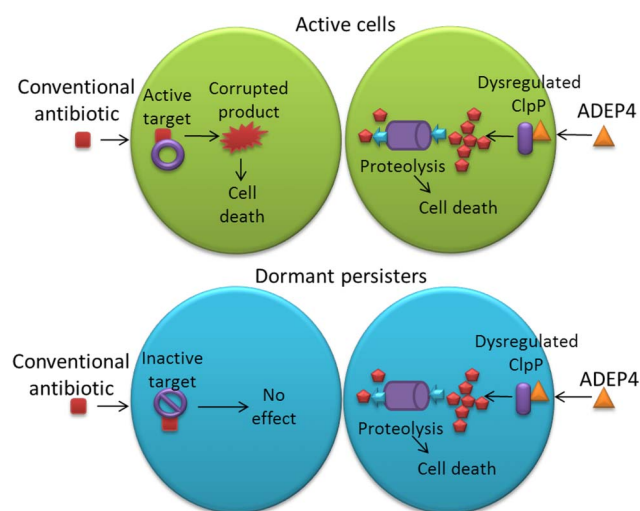


Extended Data Figure 2 | ADEP4 resistant strains are heat sensitive. Wild-type *S. aureus* ATCC 33591 and 9 ADEP4 resistant isolates with mutations in *clpP* were grown for 20 h in MHB at 44 °C in 96-well polystyrene plates. Data are representative of 3 independent experiments. Error bars represent s.d.



Extended Data Figure 3 | ADEP4 with rifampicin eradicates *S. aureus* in a hollow-fibre infection model. Antibiotics were delivered at concentrations mimicking human dosing, while the concentration of ADEP was varied over

time to match the pharmacokinetics in the mouse model. Data are representative of 3 independent experiments. Error bars represent s.d.



Extended Data Figure 4 | Conventional bactericidal antibiotics target active processes in bacterial cells (green) resulting in death. In a dormant persister (blue), the antibiotic binds an inactive target, producing no effect. ADEP4 activates and dysregulates ClpP in growing cells and in dormant persisters, resulting in eradication of the bacterial population.

DNMT1-interacting RNAs block gene-specific DNA methylation

Annalisa Di Ruscio^{1,2,3*}, Alexander K. Ebralidze^{1,2*}, Touati Benoukraf⁴, Giovanni Amabile^{1,2}, Loyal A. Goff^{5,6,7}, Jolyon Terragni⁸, Maria Eugenia Figueroa⁹, Lorena Lobo De Figueiredo Pontes^{1,2}, Meritxell Alberich-Jorda^{1,2,10}, Pu Zhang¹, Mengchu Wu⁴, Francesco D'Alò³, Ari Melnick¹¹, Giuseppe Leone³, Konstantin K. Ebralidze², Sriharsa Pradhan⁸, John L. Rinn^{1,2,5,6} & Daniel G. Tenen^{1,4}

DNA methylation was first described almost a century ago; however, the rules governing its establishment and maintenance remain elusive. Here we present data demonstrating that active transcription regulates levels of genomic methylation. We identify a novel RNA arising from the *CEBPA* gene locus that is critical in regulating the local DNA methylation profile. This RNA binds to DNMT1 and prevents *CEBPA* gene locus methylation. Deep sequencing of transcripts associated with DNMT1 combined with genome-scale methylation and expression profiling extend the generality of this finding to numerous gene loci. Collectively, these results delineate the nature of DNMT1–RNA interactions and suggest strategies for gene-selective demethylation of therapeutic targets in human diseases.

DNA methylation is a key epigenetic signature implicated in transcriptional regulation, genomic imprinting, and silencing of repetitive DNA elements^{1,2} that occurs predominantly within CpG dinucleotides. CpG dinucleotides are underrepresented in the mammalian genome (~1%) and tend to cluster within CpG islands located in the vicinity of the transcription start sites (TSSs) of the majority (~70%) of human protein-coding genes³. Although the bulk of genome is methylated at 70–80% of its CpGs, CpG islands are mostly unmethylated in somatic cells^{3,4}. This modification is mediated by the members of the DNA methyltransferase (DNMT) family, conventionally classified as *de novo* (DNMT3a and DNMT3b) and maintenance (DNMT1). In terms of epigenetic inheritance, DNMT1 has the unique ability of identifying the hemimethylated portion of newly replicated DNA. This feature may explain how DNMT1-mediated methylation could be an epigenetic mechanism maintaining the status quo. However, it certainly does not explain how DNA methylation is altered, particularly in disease states.

To examine how transcription may regulate the levels of genomic methylation, we investigated methylation dynamics of the well-studied methylation-sensitive gene *CEBPA*^{5–7}, including the potential involvement of non-coding RNAs (ncRNAs) originating within the *CEBPA* locus. Recent discoveries of functional ncRNAs have provided new regulatory clues to the control of epigenetic marks. In particular, long ncRNAs have been shown to regulate gene expression by interacting with chromatin modifiers, modulating transcription factor activity and competing for microRNA binding^{8–16}. One unexplored aspect of the regulation of gene locus DNA methylation was the possible involvement of transcripts encoded within the region.

We identified a functional RNA arising from the *CEBPA* locus, *ecCEBPA*, that regulates *CEBPA* methylation. This RNA interacts with DNMT1, resulting in prevention of *CEBPA* gene methylation and robust *CEBPA* messenger RNA production. We show that such functional

DNMT1–RNA association occurs at numerous gene loci. We thus propose a novel regulatory mechanism of gene methylation governed by RNAs.

Characterization of *ecCEBPA*

Non-coding transcripts arising from the promoter and the downstream regions of coding genes can affect the expression of the corresponding genes^{17–19}. We searched and identified transcripts upstream and downstream of the intronless *CEBPA* gene. Strand-specific reverse transcriptase PCR (RT-PCR; data not shown) and northern blot analysis of RNAs from four leukaemic cell lines, probing the region immediately after the *CEBPA* polyadenylation site, revealed the presence of a major band of ~4.5 kilobases (kb) in HL-60 and U937 (in which *CEBPA* is expressed), but not in K562 or Jurkat (in which *CEBPA* is expressed at low or undetectable levels), cell lines (Fig. 1a, b). The identified transcript is distinct from the ~2.6 kb signal, detected with a *CEBPA* coding-region probe, and correlates with *CEBPA* mRNA expression. Unlike polyadenylated *CEBPA* mRNA (Fig. 1c), this non-polyadenylated transcript is enriched in the nuclear fraction (Supplementary Fig. 1a, b), suggesting functional roles independent of protein-coding potential.

We termed this nuclear non-polyadenylated *CEBPA* ncRNA *extra-coding CEBPA* (*ecCEBPA*), as it encompasses the entire mRNA sequence in the same-sense orientation (shown by primer extension and 5' and 3' rapid amplification of complementary DNA ends (RACE); Supplementary Information and Supplementary Fig. 1c, d). Quantitative (q)RT-PCR analysis confirmed concordant expression between extra-coding and coding transcripts, in both cellular and nuclear RNAs (Fig. 1d, e). Similar correlation was observed in all tested human tissues (Supplementary Fig. 1e). Notably, *ecCEBPA* synthesis precedes the expression of its overlapping mRNA in the S phase (Supplementary Information and Supplementary Fig. 1f, g) and is regulated by both RNA polymerase (RNAP) II and III (Supplementary Information and Supplementary Fig. 1h–p), as described for other loci^{20–22}.

¹Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA. ³Università Cattolica del Sacro Cuore, Institute of Hematology, L.go A. Gemelli 8, Rome 00168, Italy. ⁴Cancer Science Institute, National University of Singapore, 117599, Singapore. ⁵Department of Stem Cell and Regenerative Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. ⁶Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁷Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. ⁸New England Biolabs, 240 County Road, Ipswich, Massachusetts 01938-2723, USA. ⁹University of Michigan, Department of Pathology, Ann Arbor, Michigan 48109-2200, USA. ¹⁰Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, 142 20 Prague, Czech Republic. ¹¹Department of Medicine, Hematology-Oncology, C-620 Weill Cornell Medical College, 1300 York Avenue, New York, New York 10021, USA.

*These authors contributed equally to this work.

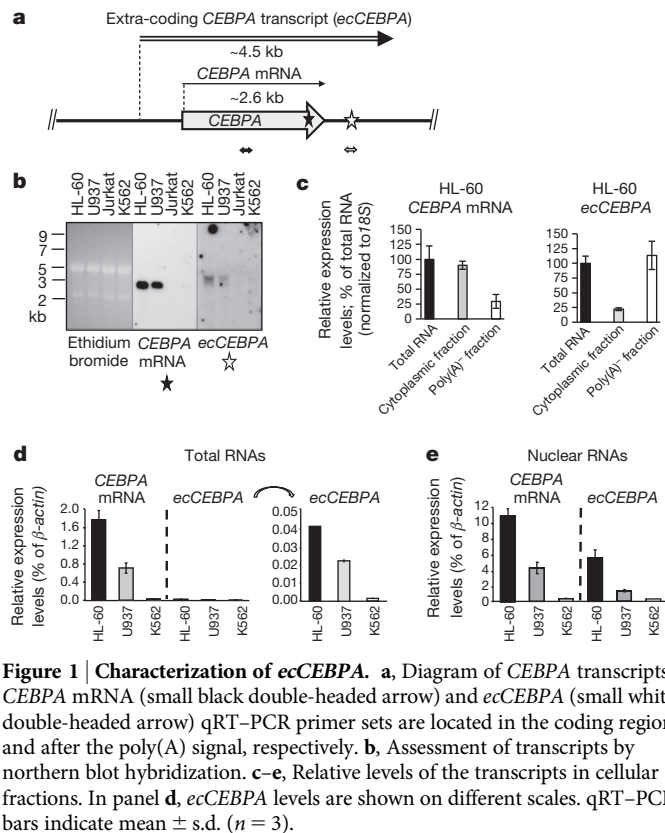


Figure 1 | Characterization of *ecCEBPA*. **a**, Diagram of *CEBPA* transcripts. *CEBPA* mRNA (small black double-headed arrow) and *ecCEBPA* (small white double-headed arrow) qRT-PCR primer sets are located in the coding region and after the poly(A) signal, respectively. **b**, Assessment of transcripts by northern blot hybridization. **c–e**, Relative levels of the transcripts in cellular fractions. In panel **d**, *ecCEBPA* levels are shown on different scales. qRT-PCR bars indicate mean \pm s.d. ($n = 3$).

ecCEBPA blocks methylation and maintains *CEBPA* mRNA

To examine the functional role of *ecCEBPA* in the regulation of *CEBPA* transcription, we performed both loss- and gain-of-function experiments. Knockdown of *ecCEBPA* in a U937 cell line (up to a fourfold decrease) achieved by short hairpin RNAs (shRNAs) targeting *ecCEBPA* (but not *CEBPA* mRNA) led to a decrease of *CEBPA* mRNA expression of similar magnitude (Fig. 2a, b), suggesting that *ecCEBPA* may regulate *CEBPA* expression. Silencing of the *CEBPA* gene can be associated with DNA methylation of the promoter^{6,7,23}. To examine whether there was a connection between *ecCEBPA* and methylation of the *CEBPA* locus, we analysed methylation within the distal promoter (located at -0.8 to -0.6 kb from the *CEBPA* TSS; Fig. 2a). Intriguingly, *ecCEBPA* knockdown led to a significant increase in DNA methylation compared to the non-targeting control (Fig. 2c and Supplementary Fig. 2a).

To investigate whether enforced expression of *ecCEBPA* was sufficient to inhibit DNA methylation, the downstream region of *ecCEBPA* (R1; Fig. 2a) was overexpressed in K562 cells expressing *ecCEBPA* and *CEBPA* mRNA at low-to-undetectable levels (Fig. 1b, d). Overexpression of only part of *ecCEBPA* was dictated by the necessity to distinguish the methylation pattern of the endogenous *CEBPA* locus from that of the ectopically expressed construct.

Ectopic expression of *ecCEBPA* R1 resulted in greater-than-threefold increase in mRNA expression (Fig. 2d), whereas overexpression of an unrelated region (located 45 kb downstream) and regions immediately outside of the *ecCEBPA* boundaries did not affect mRNA levels (Supplementary Fig. 2b–d). Moreover, a concomitant decrease of DNA methylation in three tested regions within the *CEBPA* gene, distal promoter, coding region and 3' untranslated region accompanied overexpression of *ecCEBPA* but not of the unrelated region (Fig. 2e and Supplementary Fig. 2e, f). Interestingly, comparative analysis of DNA methylation changes imposed by *ecCEBPA* overexpression versus the hypomethylating agent 5'-azacytidine (5-aza-CR), together with genome-scale analysis (reduced representation bisulphite sequencing; RRBS²⁴)

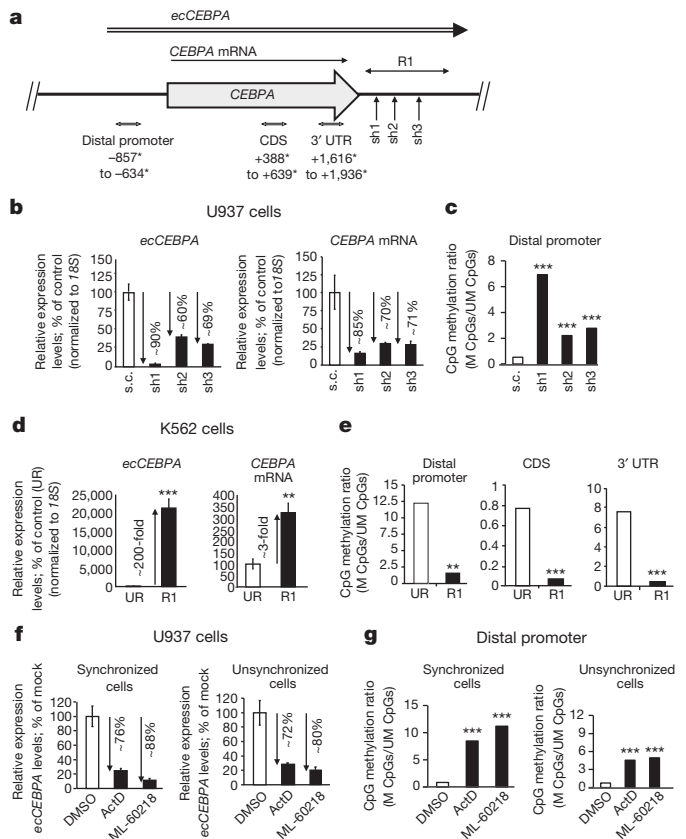


Figure 2 | Loss- and gain-of-function studies demonstrate that *ecCEBPA* maintains *CEBPA* expression by regulating methylation of the *CEBPA* locus. **a**, Diagram indicating the position of target sequences for shRNA constructs (sh1–sh3); the fragment derived from *ecCEBPA* used for overexpression (R1); regions analysed for changes in DNA methylation (distal promoter; coding sequence (CDS) and 3' untranslated region (UTR)). Asterisks indicate number of base pairs away from the *CEBPA* TSS. **b**, **c**, The results of *ecCEBPA* loss-of-function in *CEBPA*-expressing U937 cells. Effect of *ecCEBPA*-targeting shRNAs on *CEBPA* mRNA levels (qRT-PCR, bars indicate mean \pm s.d. ($n = 3$)) and methylation of the *CEBPA* promoter (**c**). DNA methylation changes are shown as the ratios of methylated (M) to unmethylated (UM) CpGs in all clones analysed per each construct ($n = 14$). s.c., scrambled control. **d**, **e**, The results of *ecCEBPA* gain-of-function studies in K562 cells, in which *CEBPA* is methylated and silenced. **d**, Effect of *ecCEBPA* upregulation on *CEBPA* mRNA levels. UR, unrelated region. qRT-PCR, bars indicate mean \pm s.d. ($n = 4$). **e**, Effect of *ecCEBPA* upregulation on methylation of the *CEBPA* locus (DNA methylation changes were assessed as described in **c** ($n = 14$ for distal promoter and $n = 6$ for CDS and 3' UTR)). **f**, **g**, The results of transcription inhibition in U937 cells. **f**, *ecCEBPA* expression levels after treatment with actinomycin D (actD) and ML-60218 in synchronized and unsynchronized cells. qRT-PCR, bars indicate mean \pm s.d. ($n = 3$). **g**, DNA methylation changes after treatment with actinomycin D and ML-60218 in synchronized ($n = 12$) and unsynchronized ($n = 10$) cells (assessed as described in **c**). DMSO, dimethylsulphoxide. Drug concentrations: actinomycin D, $0.8 \mu\text{M}$; ML-60218, $150 \mu\text{M}$. Duration of treatment was 7 h. All bisulphite sequenced clones were analysed by Fisher's exact test. $^{**}P < 0.01$; $^{***}P < 0.001$.

of DNA methylation changes imposed by *ecCEBPA* versus unrelated region overexpression, revealed that *ecCEBPA*-mediated demethylation was relatively selective to the *CEBPA* locus (Supplementary Fig. 2g–l). Indeed, increased mRNA expression and changes in methylation status within the loci of the neighbouring *CEBPG* and distant *TP73* (on chromosome 1p36) genes were achieved after 5-aza-CR treatment but not after *ecCEBPA* overexpression (Supplementary Fig. 2g–k). Furthermore, RRBS analysis of promoter and first exon regions revealed that only $\sim 3.3\%$ of the interrogated loci (396 out of 11,844)

were hypomethylated at levels similar to the *CEBPA* locus (Supplementary Fig. 2l).

Furthermore, *ecCEBPA* downregulation by the universal RNAP inhibitor actinomycin D and RNAP III-specific inhibitor ML-60218 (Supplementary Fig. 2m) led to a corresponding increase in methylation of the *CEBPA* locus, in synchronized and unsynchronized U937 cells (Fig. 2f, g and Supplementary Fig. 2n). Despite comparable decreases in *ecCEBPA* levels in both synchronized and unsynchronized cells (Fig. 2f), DNA methylation increase was more prominent in synchronized cells (Fig. 2g), suggesting a cell-cycle-specific action of the *ecCEBPA*. A similar effect was observed in ML-60218-treated HL-60 cells (Supplementary Figs 1i and 2o).

Consistently, we observed an inverse correlation between the *CEBPA* gene locus methylation and the levels of *ecCEBPA* in HL-60, U937 and K562 cell lines (Supplementary Fig. 2p).

Collectively, these data highlight the regulatory role of *ecCEBPA* in *CEBPA* gene locus methylation, most prominently during the S phase.

DNMT1 binds to RNA with greater affinity than to DNA

The changes in *CEBPA* methylation mediated by *ecCEBPA* prompted us to try to determine the mechanism through which it is achieved. Among DNMTs, it is DNMT1 whose expression and enzymatic activity peaks during S phase²⁵. Increased *ecCEBPA* expression occurs during the S phase (Supplementary Fig. 1g), whereas inhibition of *ecCEBPA* during S phase results in a substantial increase of *CEBPA* locus DNA methylation (Fig. 2f, g and Supplementary Fig. 2n). We therefore asked whether the presence of *ecCEBPA* during S phase led to RNA interference of DNMT1 activity.

To determine whether DNMT1 physically associates with *ecCEBPA* we performed RNA immunoprecipitation (RIP) with specific anti-DNMT1 antibody (Supplementary Fig. 3a). We observed *ecCEBPA* enrichment in DNMT1–RNA precipitates, demonstrating a physical interaction between *ecCEBPA* and DNMT1 (Fig. 3a, b and Supplementary Fig. 3a). Analysis of polyadenylated (poly(A)⁺) and non-polyadenylated (poly(A)[−]) fractions in DNMT1–RNA precipitates revealed enrichment of *CEBPA* transcripts in the poly(A)[−] fraction (Supplementary Fig. 3b, c), suggesting that the major component of *CEBPA* transcripts in DNMT1–RNA precipitates was *ecCEBPA*.

To investigate the molecular properties of RNA–DNMT1 interaction *in vitro*, we performed (RNA) electrophoresis mobility shift assays ((R)EMSA). RNA oligonucleotides corresponding to the 5' and 3' parts of *ecCEBPA* were selected by: (1) the ability (R2, R5 and R6) and inability (R4) to fold into stem-loop structures²⁶; and (2) the presence (R2, R5 and R6) or absence (R4) of CpG dinucleotides (Fig. 3a). RNA–DNMT1 complex formation was observed with all RNA oligonucleotides able to fold into stem-loop structures (Fig. 3c–e). Unlike DNA²⁷, CpG to UpG substitutions, neutral with regard to secondary structures, did not affect binding (mutR2; Fig. 3c). By contrast, mutations abrogating RNA folding ability affected RNA–DNMT1 binding (mut R5; Fig. 3e). Analyses extended to a number of RNA oligonucleotides not related to *ecCEBPA* (single-stranded R1 and R3 and double-stranded R13; Supplementary Fig. 3d) confirmed DNMT1 binding to stem-loop-structured RNAs (Supplementary Information and Supplementary Fig. 3e, f). Importantly, REMSAs performed in the presence of increasing concentrations of spermine, a molecule with four positive charges at high density, excluded a case of charge–charge interactions (Supplementary Fig. 3g), supporting a strong element of structural recognition between DNMT1 and RNA.

To determine the relative affinity of DNMT1 for *ecCEBPA* versus DNA, single-stranded RNA oligonucleotides capable of forming secondary structures (R5) and corresponding unmethylated double-stranded DNA (umDNA; D5/D6), hemimethylated double-stranded DNA (hmDNA; D5/D6) and fully methylated double-stranded DNA (mDNA; D5/D6) (Fig. 3a), at a constant molar concentration, were titrated with an increasing range of DNMT1 enzyme concentrations using EMSA. RNA formed complexes beginning at <0.013 μ M

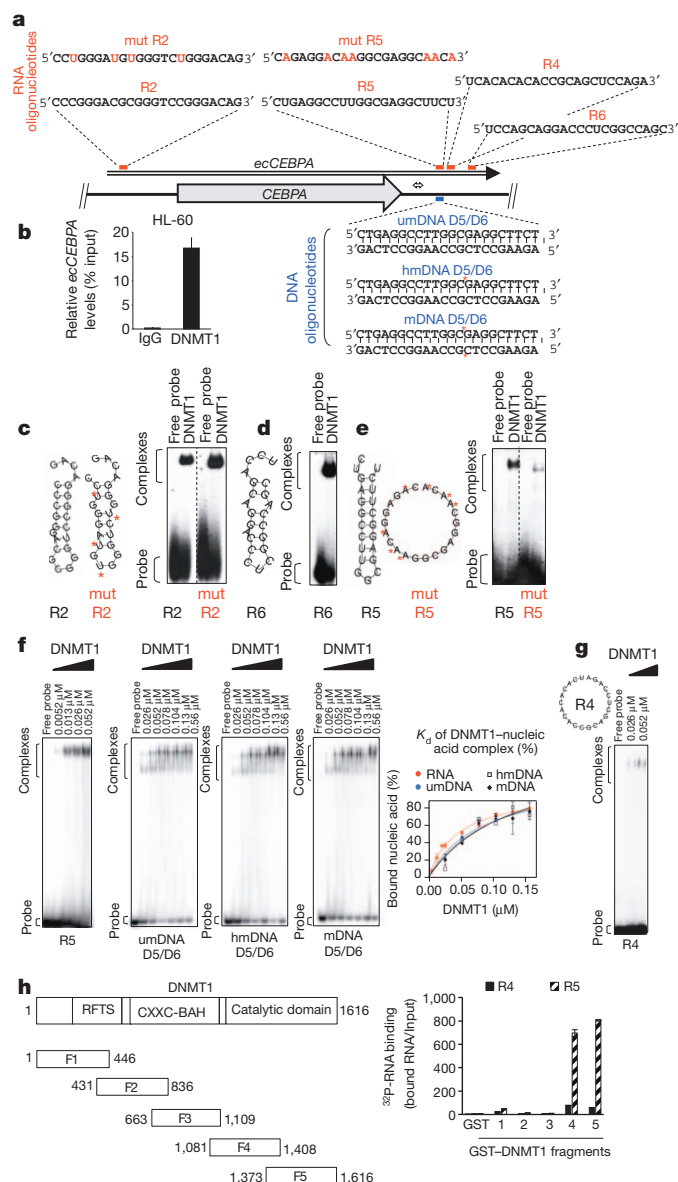


Figure 3 | *ecCEBPA*–DNMT1 interactions: DNMT1 binds to RNA with greater affinity than to DNA. **a**, Diagram showing position of qRT–PCR primers used in RIP (double-headed arrow) and RNA and DNA oligonucleotides used in EMSA and REMSA. Asterisks indicate position of methylated cytosines. **b**, *ecCEBPA* is immunoprecipitated with anti-DNMT1 antibody. qRT–PCR, bars indicate mean \pm s.d. **c**, RNA–DNMT1 binding is not affected by the absence of CpG dinucleotides (right). Left and middle: RNA oligonucleotide R2 and its mutated form mut R2 (asterisks indicate cytosines substituted to uridines), both able to form stem-loop structures. **d**, RNA oligonucleotides able to form stem-loop structure bind DNMT1 (R6). **e**, R5 RNA oligonucleotide forming stem-loop structure (R5) has a greater DNMT1 affinity compared to mut R5, which is unable to fold into stem-loop (taken in equimolar amounts) at the same DNMT1 concentration. **f**, Left four panels, REMSA and EMSA performed with the fixed concentration of single-stranded RNA and double-stranded DNA oligonucleotides (1 nM) and increasing concentrations of DNMT1 protein. Right, nonlinear regression analysis of bound RNA/DNA versus DNMT1 concentrations. Error bars indicate s.d. from two independent experiments. **g**, REMSA showing that RNA oligonucleotide R4, which is unable to form stem-loop structure, displays lower DNMT1 affinity as compared to R5 (f, left) at the same DNMT1 concentrations. **h**, Left, schematic diagram showing the DNMT1 domains and the GST–DNMT1-isolated fragments (F1–F5). BAH, bromo-adjacent homology; RFTS, replication foci targeting sequence. Right, GST–DNMT1 pull-down assay demonstrating binding of the folded RNA oligonucleotide R5 to the catalytic domain of DNMT1.

DNMT1 and DNA at $>0.026\ \mu\text{M}$ DNMT1 (Fig. 3f), with a mean dissociation constant (K_d) for RNA of $0.045\ \mu\text{M}$ and between 0.082 and $0.14\ \mu\text{M}$ for DNA, indicating that RNA has a stronger affinity for the enzyme than DNA (Fig. 3f). Consistently, RNA unable to fold into stem-loop structures (R4; Fig. 3g) did not display the same affinity for DNMT1 as 'folded' RNA (R5; Fig. 3f, g), demonstrating that RNA secondary structure represents an essential feature of RNA–DNMT1 complex formation.

Finally, to assess which DNMT1 domain is required for the RNA binding, DNMT1–glutathione S-transferase (GST)-purified domains (Fig. 3h) were incubated with RNA oligonucleotides able or unable to fold into stem-loop structures (R5 and R4, respectively; Fig. 3e, g). The catalytic domain, including the target recognition domain²⁸ shared by both fragments F4 and F5, selectively bound the 'folded' RNA oligonucleotide (Fig. 3h). Next, we deleted the DNMT1 region including the sequence overlapping F4 and F5. Unfortunately, even minimal removal of the target recognition domain led to disruption of DNMT1 enzymatic activity (data not shown), making further refinement of the binding domain unfeasible.

Collectively, these data indicate that RNA can associate with DNMT1. This interaction is not contingent upon the presence of CpG dinucleotides, is not a trivial ion pairing, and is dependent upon certain RNA secondary-structure features. Importantly, DNMT1, through its catalytic domain, binds with higher affinity to folded RNA than to DNA.

Transcription interferes with DNMT1 activity

To examine whether newly synthesized transcripts could interfere with the ability of DNMT1 to methylate hmDNA, we performed a combined *in vitro* transcription–DNA methylation assay. A hmDNA segment (bottom-strand methylated) was engineered downstream of the T7 RNAP promoter (Supplementary Fig. 4a–h) and DNMT1 methylase activity was monitored in the presence and absence of transcription (Fig. 4a–d). In the absence of polymerase there was, as expected, increased DNA methylation of the upper strand (Fig. 4d–f and Supplementary Fig. 4i–j). By contrast, no changes in DNA methylation were observed in the presence of both polymerases and DNMT1 (Fig. 4c–f and Supplementary Fig. 4i–j). Standard *in vitro* DNA methylation assays confirmed the enzymatic impairment of DNMT1 mediated by ribo-oligonucleotides (Fig. 4g). Similarly, T7 RNA polymerase²⁹-induced transcription in living cells led to a pronounced decrease in DNA methylation (Supplementary Information and Supplementary Fig. 4k–p).

Thus, RNA can complex with and affect DNMT enzymatic activity *in vitro*^{30–32} and in living cells. These findings suggest that RNAs arising from methylation-sensitive genes and their promoters can regulate expression of the corresponding genes by interfering with DNA methylation.

RNA inhibition of DNA methylation is a global effect

Our observations suggested an inverse correlation between RNA–DNMT1 complexes and methylation of the *CEBPA* locus. Therefore, we sought to explore the extent of DNMT1–RNA association in other genomic loci with respect to DNA methylation and gene expression profiles. Complementary DNA libraries made of RNAs coimmunoprecipitated with anti-DNMT1 antibody (DNMT1 library) and IgG (control library) were tested for *ecCEBPA* enrichment ('quality control'; Supplementary Fig. 5a) and subsequently analysed by massively parallel sequencing¹¹. Using 76-base paired-end sequencing, we produced a total of 30.25 and 26.95 million pair reads for DNMT1 and control libraries, respectively (detailed analysis described in Methods). All significant DNMT1 peaks (a total of 16,186; $P < 0.0001$; false discovery rate of 7.5%) were annotated with CEAS³³ build on RefSeq hg19 (Supplementary Fig. 5b). All DNMT1 peaks were also annotated using the known RNAs databases provided by HOMER³⁴ (Supplementary Fig. 5c). We focused on genomic regions encompassing the 3 kb upstream of the TSS and downstream to the transcription ending site of the annotated

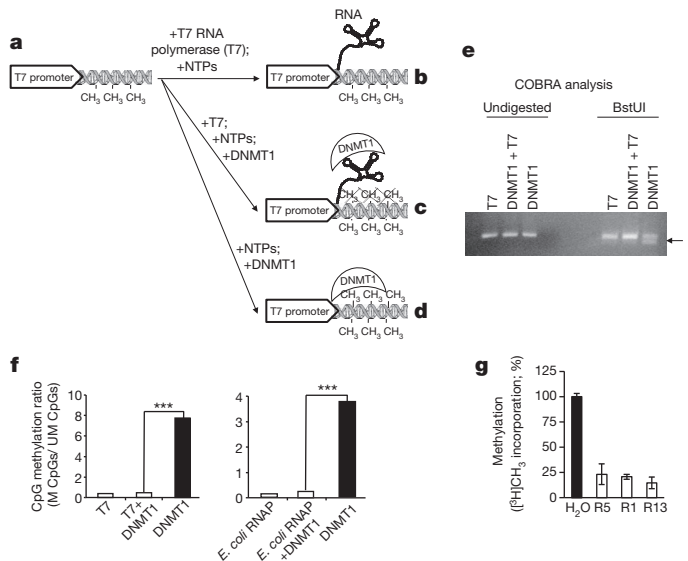


Figure 4 | Transcription impedes DNA methylation. **a–d**, Diagram showing the parallel *in vitro* transcription–methylation assays performed on a hemimethylated template containing the T7 promoter (Supplementary Fig. 4) with and without combinations of RNA polymerase, DNMT1, or both. NTPs, nucleotide triphosphates. **e**, DNMT1 exerts enzymatic activity only in the absence of transcription. Combined bisulphite restriction analysis assay (COBRA) analysis of methylation patterns acquired in reactions shown in **b–d**. **f**, DNA methylation changes are shown as the ratios of methylated to unmethylated CpGs in all clones analysed per construct ($n = 5$). The same effect was observed with two different RNA polymerases: T7 and sigma-saturated ($\sigma 70$)-holoenzyme (*Escherichia coli* RNAP). DNA methylation changes were analysed by Fisher's exact test ($***P < 0.001$). **g**, *In vitro* DNMT1 assay demonstrating DNMT1 enzymatic impairment by RNA oligonucleotides. The assay was performed using *ecCEBPA*-related and -unrelated RNA oligonucleotides. Sequences and position of the ribo-oligonucleotides are shown in Fig. 3a and Supplementary Fig. 3. Error bars indicate mean \pm s.d. ($n = 2$).

genes, referred to as 'gene loci'. We identified 6,042 gene loci containing one or more peaks from the DNMT1 library (Methods).

To confirm that DNMT1 RIP-seq peaks were associated with actual transcribed elements, RNA-seq was conducted on poly(A)[−] HL-60 RNA. In total, 375 million 76-bp paired-end reads were aligned to hg19 using TopHat²⁵ and assembled using Cufflinks³⁶, and 14,077 (87.02%) of the specific DNMT1 peaks overlapped with a transcribed element from the RNA-seq assembly of the poly(A)[−] HL-60 RNA fraction. Thus, the vast majority of DNMT1-interacting RNAs (DiRs) were not polyadenylated.

In addition, we performed a similar analysis with total HL-60 RNA (300 million 76-bp paired-end reads). In total, 14,497 specific DNMT1 peaks (89.61%) were found to overlap with transcripts from the total HL-60 RNA-seq assembly. A merged assembly of the two RNA libraries validated a total of 15,238 (94.20%) DNMT1 RIP-seq peaks (Fig. 5a). These findings confirmed the existence of DiRs on a genome-wide level. Next, we assessed the linkage between genomic loci giving rise to DiRs, levels of genomic methylation by RRBS²⁴ and expression of the corresponding nearby genes by microarray analysis, performed on HL-60 cells. Within all 15,806 RRBS-covered loci, 10,973 loci were not covered by DNMT1-specific peaks (DNMT1-unbound group) and 4,833 loci were covered by DNMT1-specific peaks (DNMT1-bound group). These 4,833 loci represent the majority (79.99%) of all 6,042 gene loci identified by DNMT1 RIP-seq (Supplementary Fig. 5d).

Within DNMT1-bound and -unbound groups, genes were stratified according to expression and methylation levels (the latter computed as the mean of all CpG β -scores from -2 kb from the TSS to the end of the first intron). A negative correlation between DNMT1–RNA

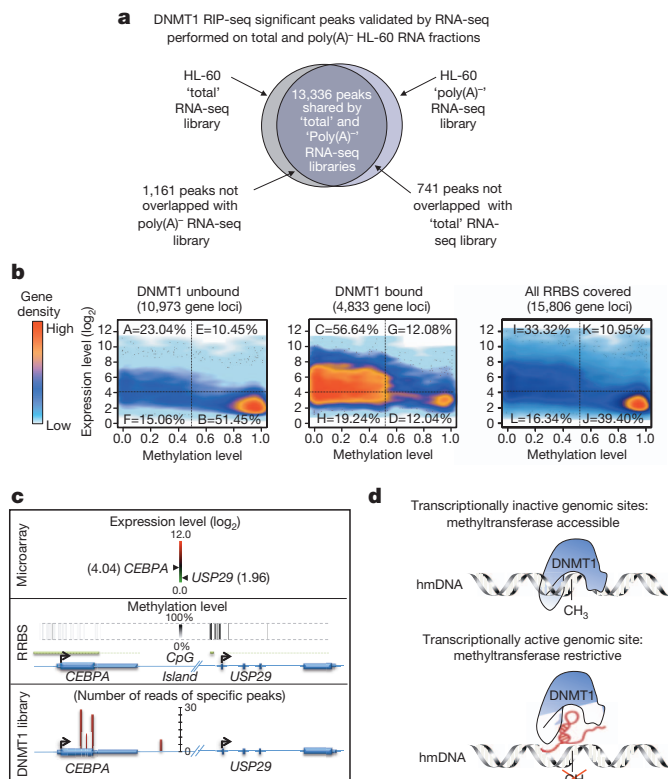


Figure 5 | Genome-wide alignment of DNMT1-bound and -unbound transcripts, DNA methylation and gene expression. **a**, Two-way Venn diagram showing DNMT1-specific peaks overlapping with transcribed elements identified in HL-60 total and poly(A)⁺-depleted RNA-seq libraries. **b**, Cloud plots representing genes within DNMT1-unbound, -bound and all RRBS-covered groups stratified by DNA methylation and expression levels. All genes are presented in Supplementary Data 2. **c**, Examples of genes from the C (*CEBPA*) and B (*USP29*) clusters. Peaks are visualized using the SSRs (site identification from short sequence reads)⁴⁸. **d**, Model of DNMT1 sequestration. Top, DNMT1 can access transcriptionally inactive hemimethylated genomic regions. Bottom, DNMT1 cannot access transcriptionally active hemimethylated genomic regions.

association with gene locus methylation status was observed (Supplementary Fig. 5e).

Next, we clustered genes within both groups according to levels of expression and methylation. We defined genes as 'expressed' or 'low or not expressed' if the log₂ score was above or below 4, respectively; and 'hypomethylated' or 'methylated' if the mean of all CpG scores was below or above 50%, respectively (Methods). This approach allowed us to identify four clusters within DNMT1-unbound, DNMT1-bound, and all RRBS-covered groups (Fig. 5b). Hypomethylated and expressed genes appeared to be predominant in the DNMT1-bound group (cluster C), accounting for 56.64%, whereas hypermethylated and low or unexpressed genes represented the 51.45% in the DNMT1-unbound group (cluster B). Moreover, the numbers of genes in clusters B (5,646 genes) and C (2,737 genes) were significantly higher than numbers of genes in clusters A, F, E (2,528, 1,653, 1,146) and G, H, D (584, 930, 582), respectively ($P < 0.0001$). Examples of genes from clusters B and C are presented in Fig. 5c and Supplementary Fig. 5f–h). Furthermore, genes from cluster C belonged to a multiplicity of biological processes, indicating the diversity of DiRs (Supplementary Fig. 6a). Interestingly, ~60% of these Biological Process Gene Ontology (BP-GO) terms ($P \leq 0.01$) were shared with pyknon (non-random pattern of repeated elements)-related BP-GO³⁷. This overlap is 71-fold higher than expected. Moreover, among all DNMT1 RIP-seq peaks, 46% carry at least one pyknon (Supplementary Fig. 6b) suggesting a potential relation between DiRs and pyknons.

Grouping of genes in clusters A, F, E, G, H and D could result from technical limitations of RRBS, contingent upon the genomic location of the restriction sites and the DNA library size selection³⁸, or these genes may be governed by yet another mechanism of transcriptional control.

In conclusion, we have generated the first comprehensive map cross-referencing DiRs to DNA methylation and gene expression. These data demonstrate that RNA–DNMT1 association is widespread and might modulate genomic DNA methylation (Fig. 5d).

Discussion

This study explores the role of a new class of RNAs: DNMT1-interacting RNAs. Using the *CEBPA* gene as a model, we show that mRNA transcription is accompanied by the production of an additional RNA species, *ecCEBPA*. In every instance studied, DNA methylation levels are inversely correlated with *ecCEBPA* levels, and the extent of DNA methylation is determined by the absence or presence of *ecCEBPA* (Fig. 2).

We demonstrate that *ecCEBPA* associates with DNMT1 and establish a functional link between *ecCEBPA* and *CEBPA* expression as through RNA–DNMT1 association.

We show that RNAs capable of adopting stem-loop structures exhibit the potential to associate with DNMT1, suggesting that the basis of this preferential interaction is recognition of RNA secondary structure (Fig. 3).

Importantly, we demonstrate that this type of RNA–DNMT1 association is not restricted to the *CEBPA* gene locus. We have globally identified RNA species associated with DNMT1 and their relationship to DNA methylation and gene expression. These alignments defined a large set of expressed unmethylated genes and a complementary set of silent methylated genes that could possibly be induced following expression of the DiR.

Our findings suggest a model of site-specific DNMT1 sequestration in which RNAs act as a shield, halting DNMT1 and thus modulating DNA genomic methylation at their site of transcription (Fig. 5d). Indeed, the loss of *CEBPA* locus methylation following overexpression of *ecCEBPA* does not support a model of trivial titration ('squelching'³⁹) of DNMT1 but suggests a *cis*-regulatory role of the DiR. We propose a model wherein RNAs contain a locus-selective triplex/quadruplex⁴⁰-forming part, the 'anchor', mooring the DNMT1–RNA complex to the locus, and a DNMT1-interacting part, the 'bait', a stem-loop-like-forming sequence serving to lure the DNMT1 into association. DNMT1 sequestration by RNA does resemble a competing mechanism described for other regulatory RNAs, for example, competing endogenous RNAs^{15,16,41}. However, unlike competing endogenous RNAs, the *ecCEBPA* model also introduces the requirements for both functional and physical co-compartmentalization of the RNA, its parental locus, and DNMT1. Given the ability of DiRs to bind DNMT1, it is tempting to suggest that DiRs represent a novel class of RNA regulons⁴².

Taken together, these data support the hypothesis that RNA participates in the establishment of genomic methylation patterns by interacting with DNMT1 and pave the way for the site-specific adjustments of aberrant DNA methylation.

METHODS SUMMARY

RIP-seq. Double-stranded cDNA from total RNA immunoprecipitated with DNMT1 antibody (Abcam) or IgG (Sigma Aldrich) was synthesized using the Just cDNA Double-Stranded cDNA Synthesis Kit (Agilent Technology) according to the manufacturer's instructions. cDNA libraries were paired-end sequenced on an Illumina GA IIx.

MassARRAY and RRBS. MassARRAY and RRBS were performed as described^{13,44}. For RRBS, sequenced reads were mapped to the reference genome hg19 using RRBSmap⁴⁵ allowing two mismatches and gene loci were scored by Genome Bisulfite Sequencing Analyzer (GBSA)⁴⁶. Differentially methylated domains were computed using the R/MethylKit package⁴⁷.

RNA-seq. Total and non-polyadenylated RNA were depleted of ribosomal RNA with a Ribo-Zero™ Magnetic Gold Kit. Double-stranded cDNA libraries were

constructed using ScriptSeq v2 RNA-Seq Library Preparation Kit. Libraries were sequenced (1 per lane) on a Hi-Seq-2000 Illumina instrument.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 November 2011; accepted 21 August 2013.

Published online 9 October 2013.

- Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nature Rev. Genet.* **3**, 415–428 (2002).
- Robertson, K. D. DNA methylation, methyltransferases, and cancer. *Oncogene* **20**, 3139–3155 (2001).
- Illingworth, R. S. *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **6**, e1001134 (2010).
- Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA* **103**, 1412–1417 (2006).
- Tenen, D. G. Disruption of differentiation in human cancer: AML shows the way. *Nature Rev. Cancer* **3**, 89–101 (2003).
- Tada, Y. *et al.* Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein α activity in lung cancer. *J. Natl. Cancer Inst.* **98**, 396–406 (2006).
- Hackanson, B. *et al.* Epigenetic modification of CCAAT/enhancer binding protein α expression in acute myeloid leukemia. *Cancer Res.* **68**, 3142–3151 (2008).
- Gupta, R. A. *et al.* Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076 (2010).
- Huarte, M. *et al.* A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* **142**, 409–419 (2010).
- Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
- Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
- Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322**, 1717–1720 (2008).
- Guttman, M. & Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346 (2012).
- Ebralidze, A., Wang, Y., Petkova, V., Ebralidze, K. & Junghans, R. P. RNA leaching of transcription factors disrupts transcription in myotonic dystrophy. *Science* **303**, 383–387 (2004).
- Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
- Tay, Y. *et al.* Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147**, 344–357 (2011).
- Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. *Nature Cell Biol.* **10**, 1106–1113 (2008).
- Preker, R. *et al.* RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
- Ebralidze, A. K. *et al.* *PU.1* expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared *cis*-regulatory element. *Genes Dev.* **22**, 2085–2092 (2008).
- Oler, A. J. *et al.* Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Struct. Mol. Biol.* **17**, 620–628 (2010).
- Listerman, I., Bledau, A. S., Grishina, I. & Neugebauer, K. M. Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. *PLoS Genet.* **3**, e212 (2007).
- Raha, D. *et al.* Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc. Natl Acad. Sci. USA* **107**, 3639–3644 (2010).
- Wouters, B. J. *et al.* Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in *NOTCH1*. *Blood* **110**, 3706–3714 (2007).
- Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
- Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71**, 865–873 (1992).
- Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431 (2003).
- Pradhan, S. & Esteve, P. O. Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. *Biochemistry* **42**, 5321–5332 (2003).
- Song, J., Rechkobl, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036–1040 (2011).
- Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA* **83**, 8122–8126 (1986).
- Svedruzic, Z. M. Mammalian cytosine DNA methyltransferase Dnmt1: enzymatic mechanism, novel mechanism-based inhibitors, and RNA-directed DNA methylation. *Curr. Med. Chem.* **15**, 92–106 (2008).
- Bolden, A., Ward, C., Siedlecki, J. A. & Weissbach, A. DNA methylation. Inhibition of *de novo* and maintenance methylation *in vitro* by RNA and synthetic polynucleotides. *J. Biol. Chem.* **259**, 12437–12443 (1984).
- Bolden, A. H., Nalin, C. M., Ward, C. A., Poonian, M. S. & Weissbach, A. Primary DNA sequence determines sites of maintenance and *de novo* methylation by mammalian DNA methyltransferases. *Mol. Cell. Biol.* **6**, 1135–1140 (1986).
- Shin, H., Liu, T., Manrai, A. K. & Liu, X. S. CEAS: *cis*-regulatory element annotation system. *Bioinformatics* **25**, 2605–2606 (2009).
- Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
- Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–578 (2012).
- Tsirigos, A. & Rigoutsos, I. Human and mouse introns are linked to the same processes and functions through each genome's most frequent non-conserved motifs. *Nucleic Acids Res.* **36**, 3484–3493 (2008).
- Bock, C. *et al.* Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nature Biotechnol.* **28**, 1106–1114 (2010).
- Gill, G. & Ptashne, M. Negative effect of the transcriptional activator GAL4. *Nature* **334**, 721–724 (1988).
- Frank-Kamenetskii, M. D. & Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **64**, 65–95 (1995).
- Karreth, F. A. *et al.* *In vivo* identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* **147**, 382–395 (2011).
- Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8**, 533–543 (2007).
- Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13–27 (2010).
- Smith, Z. D., Gu, H., Bock, C., Gnirke, A. & Meissner, A. High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48**, 226–232 (2009).
- Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* **10**, 232 (2009).
- Benoukr, T., Wongphayak, S., Hadi, L. H., Wu, M. & Soong, R. GBSA: a comprehensive software for analysing whole genome bisulfite sequencing data. *Nucleic Acids Res.* **41**, e55 (2013).
- Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
- Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**, 5221–5231 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by grants CA118316, CA66996 and HL56745 from the National Institutes of Health (NIH) to D.G.T., the Italian Foundation for Cancer Research (FIRC) 'Leonino Fontana e Maria Lionello' fellowship, the NIH T32 HL007917-11A1 and the Società Italiana di Ematologia Sperimentale (SIES) 'Dr. Tito Bastianello' fellowship to A.D.R.; FAMRI CIA (103063) grant to A.K.E.; Fondazione Roma 'Progetto cellule staminali' to G.L. and F.D'A.; the American Italian Cancer Foundation Fellowship (AICF) to G.A.; Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)—grant no. 2011/11822-6—to L.L.D.F.P.; the National Research Foundation and the Singapore Ministry of Education under its Centres of Excellence initiative to M.W. and T.B.; the MSMT Navrat grant LK21307 to M.A.-J. S.P. and J.T. were supported by New England Biolabs. We thank R. White, D. Johnson, M. Frank-Kamenetskii, S. M. Mirkin, D. Gautheret, B. Tazon-Vega, C. Bonifer, C. Bock, M. T. Voso, J. J. Dunn (deceased) and R. A. M. Fouchier for helpful advice and reagents; I. Rigoutsos for providing the latest released pyknons database; all the members of the Tenen Laboratory; P. Tan and T. S. Ting from the Genome Institute of Singapore; R. Soong from the Cancer Science Institute Translational Interface; J. LaVecchio and G. Buruzula from the Harvard Stem Cell Institute/Joslin Diabetes Center flow cytometry facility; and F. Hyde from Epicentre-Illumina. This research is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research (STaR) Investigator Award (D.G.T.).

Author Contributions D.G.T. supervised the project; A.D.R., A.K.E. and D.G.T. conceived and designed the study; A.D.R., A.K.E., G.A., P.Z., M.A.-J., F.D'A., S.P., L.L.D.F.P. and J.T. performed experiments; M.W. performed sequencing and microarray experiments; T.B. and L.A.G. analysed RIP-seq, RNA-seq, RRBS and microarray data; M.E.F. and A.M. performed the MassARRAY experiment and assisted in the analysis; A.D.R., A.K.E., G.L., K.K.E., J.L.R. and D.G.T. wrote the paper.

Author Information Sequencing and microarray data sets are available for download at Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/projects/geo/>) under the following accession numbers: microarray expression, GSE32153; RIP-seq, GSE32162; RRBS, GSE32168; and RNA-seq, GSE41279. The accession number for the project is GSE32260. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.G.T. (daniel.tenen@nus.edu.sg).

METHODS

Cell culture. All cell lines were obtained from ATCC and grown according to the manufacturer's instructions in the absence of antibiotics. The DNMT1 hypomorphic HCT116 cell line and its wild-type counterpart HCT116 were grown in McCoy's 5A modified medium supplemented with 10% FCS.

RNA isolation and northern blot analysis. Total RNA isolation, electrophoresis, transfer and hybridization were carried out as described⁴⁹. Cytoplasmic RNA was isolated with the Paris kit (Ambion) according to the manufacturer's recommendations. For the preparation of nuclear RNAs we used a method derived from protocols of nuclei isolation⁵⁰, with minor modifications. In brief, equal amounts of viable cells (~50 million) were washed with ice-cold PBS supplemented with 5 mM vanadyl complex, 1 mM phenylmethylsulphonyl fluoride (PMSF) and re-suspended in the ice-cold lysis buffer: 1 × buffer A (10 mM HEPES-NaOH, pH 7.6, 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 mM Na butyrate), 1.25 M sucrose, 10% glycerol, 5 mg ml⁻¹ BSA, 0.5% NP-40, freshly supplemented with protease inhibitors (2 mM leupeptin, add as ×400; 2 mM pepstatin, add as ×400; 100 mM benzamidine, add as ×400; a protease inhibitor cocktail (Roche Applied Science), 1 tablet per 375 µl H₂O, add as ×100; 100 mM PMSF, add as ×100), 2 mM vanadyl complex (New England Biolabs) and 20 units per ml RNase inhibitor (RNAGuard; Amersham Biosciences). Samples were incubated at 0 °C for ~10 min and passed through 40 up-and-down strokes in a Dounce homogenizer (10 with pestle A and 30 with pestle B). The pelleted nuclei were re-suspended in 0.5 ml lysis buffer and diluted with 2.25 ml dilution buffer (2.13 ml 'cushion' buffer plus 0.12 ml 0.1 g ml⁻¹ BSA), freshly supplemented with protease inhibitors and overlaid onto 2 ml 'cushions' (20 ml cushion buffer consists of 15 ml double-deionized (dd)H₂O, 15 ml 20× buffer A, 30 ml glycerol, 240 ml 2.5 M sucrose; freshly supplemented with protease inhibitors) into one SW 55 Ti tube and centrifuged at 100,000g for 60 min at 4 °C. The pelleted nuclei were re-suspended in 1 ml storage buffer (1.75 ml ddH₂O, 2 ml glycerol, 0.2 ml 20× buffer A), freshly supplemented with protease inhibitors. Nuclear RNAs were extracted as described⁵⁰. All total, cytoplasmic and nuclear RNA samples used in this study were treated with DNase I (10 U of DNase I per 3 µg of total RNA; 37 °C for 1 h; in the presence of RNase inhibitor). After DNase I treatment, RNA samples were extracted with acidic phenol (pH 4.3) to eliminate any remaining traces of DNA. Polyadenylated and non-polyadenylated RNA fractions were selected with the MicroPoly(A) Purist purification kit (Ambion). cDNA syntheses were performed with Random Primers (Invitrogen) with Transcriptor Reverse Transcriptase (Roche Applied Science) according to the manufacturer's recommendation. cDNA was purified with a High Pure PCR Product Purification Kit (Roche Applied Science).

qRT-PCR. SYBR green reactions were performed using iQ Sybr Green supermix (Bio-Rad); PCR conditions: 95 °C (10 min) followed by 40 cycles of 95 °C (15 s) and 60 °C (1 min) and 72 °C (1 min). TaqMan analysis was performed using Hotstart Probe One-step qRT-PCR master mix (USB); PCR conditions: 50 °C (10 min), 95 °C (2 min), followed by 40 cycles of 95 °C (15 s) and 60 °C (60 s). Primers and probes are presented in Supplementary Methods. qRT-PCR primer set for the *CEBPA* mRNA is located in the coding region (black double-headed arrow in Fig. 1a) and after the poly(A) signal for *ecCEBPA* (white double-headed arrow Fig. 1a).

Primer extension and 5'/3' RACE. cDNA from the HL-60 cell line was synthesized as described above and run in alkaline conditions⁵¹. Southern blot transfer and hybridization with oligonucleotide AL16 were performed as reported previously⁵¹. Oligonucleotide sequences are shown in Supplementary Methods. 5'/3' RACE was performed on two myeloid cell lines—HL-60 and U937—using the Exact START Eukaryotic mRNA 5'- & 3'- RACE Kit (Epicentre) according to the manufacturer's instructions. See Supplementary Methods for primer sequences.

Double thymidine block (early S phase block). HL-60 cells were grown to 70–80% confluence, washed twice with 1 × PBS and cultured in DMEM (10% FCS) plus 2.5 mM thymidine for 18 h (first block). Thymidine was washed out with 1 × PBS and cells were grown in DMEM (10% FCS). After 8 h cells were cultured in the presence of thymidine for 18 h (second block) and then released as described. Synchrony was monitored by flow cytometry analysis (propidium iodide staining) using a LSRII flow cytometer (BD Biosciences) at the Harvard Stem Cell Institute/Beth Israel Deaconess Medical Center flow cytometry facility.

DRB, ML-60218, α-amanitin and actinomycin D treatment. After release from double thymidine block, HL-60 cells were treated with 100 µM 5,6-dichlorobenzimidazole 1-β-D-ribofuranoside⁵² (DRB; Sigma-Aldrich) for 1, 2 and 3 h. HL-60 cells were treated with 12.5, 25, 50 or 100 µM ML-60218 (refs 53, 54; Calbiochem) for 24 h. HL-60 cells were treated with 5, 23, 50, 75, 100 or 150 µg ml⁻¹ α-amanitin (Sigma-Aldrich) for 14 h. Synchronized and unsynchronized U937 cells were treated with ML-60218 at 100 µM and actinomycin D (Sigma-Aldrich) at 150 µM for 7 h. Total RNA was collected as described above and expression levels of *CEBPA*, *ecCEBPA* and 5S were measured by TaqMan qRT-PCR.

Nuclear chromatin immunoprecipitation and RIP. Chromatin immunoprecipitation (ChIP) was performed as described previously¹⁹. Fold enrichment was calculated using the formula $2^{(-\Delta\Delta Ct(\text{ChIP/non-immune serum}))}$. Antibodies used for ChIP are listed in Supplementary Table 1.

RIP was performed as described in ref. 14. Day 1: crosslinked nuclei were collected as follows: 60×10^6 HL-60 cells were crosslinked with 1% formaldehyde (HCHO; formaldehyde solution, freshly made: 50 mM HEPES-KOH, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 11% formaldehyde) for 10 min at room temperature (21 °C). Crosslinking was stopped by adding one-tenth the volume of 2.66 M glycine, kept for 5 min at room temperature and 10 min on ice. Cell pellets were washed twice with ice-cold PBS (freshly supplemented with 1 mM PMSF). Cell pellets were re-suspended in cell lysis buffer (volume 4 ml): 1 × buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 0.5% NP-40, freshly supplemented with protease inhibitors (protease inhibitors cocktail: Roche Applied Science, 1 tablet per 375 µl H₂O; add as ×100), 1 mM PMSF and 2 mM vanadyl complex (NEB). Cells were incubated at 0 °C for 10–15 min and homogenized by Dounce (10 strokes with pestle A and 40 strokes with pestle B). Nuclei were recovered by centrifugation at 2,000 r.p.m. for 10 min at 4 °C. Nuclei were re-suspended in 3 ml 1 × re-suspension buffer (50 mM HEPES-NaOH, pH 7.4, 10 mM MgCl₂) supplemented with 1 mM PMSF and 2 mM vanadyl complex. DNase treatment (250 U ml⁻¹) was performed for 30 min at 37 °C, and EDTA (final concentration 20 mM) was added to stop the reaction. Re-suspended nuclei were sonicated once for 20 s (1 pulse every 3 s) at 30% amplitude (Branson Digital Sonifier). Immunoprecipitation for RIP was performed as follows: before preclearing, the sample was adjusted to 1% Triton X-100, 0.1% sodium deoxycholate, 0.01% SDS, 140 mM NaCl, protease inhibitors, 2 mM vanadyl complex and 1 mM PMSF to facilitate solubilization. Preclearing step: ~50 µl magnetic beads (Protein A or G Magnetic Beads, NEB) were added to the sample and incubation was carried out for 1 h on a rocking platform at 4 °C. Beads were removed in a magnetic field. The sample was divided into three aliquots: (1) antibody of interest: either DNMT1 antibody (Abcam) or anti-cap antibody (Anti-m₃G-/m⁷G-cap; Synaptic Systems); (2) pre-immune serum: IgG (Sigma-Aldrich); (3) no antibody, no serum (input). 5 µg antibody or pre-immune serum was added to the respective aliquot and incubation performed on a rocking platform overnight at 4 °C. Input was stored at -20 °C after addition of SDS to 2% final concentration. Day 2: 200 µl of protein A-coated superparamagnetic beads (enough to bind 8 µg IgG) were added to the samples and incubated on a rocking platform for 1 h at 4 °C. Six washes were performed with immunoprecipitation buffer (150 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 1 mM EGTA, pH 8.0, 1% Triton X-100, 0.5% NP-40 freshly supplemented with 0.2 mM vanadyl complex and 0.2 mM PMSF) in a magnetic field. Proteinase K treatment to release DNA/RNA into solution and to reverse HCHO crosslinking was performed in 200 µl of: 100 mM Tris-HCl, pH 7.4; 0.5% SDS for the immunoprecipitated samples and in parallel for the input; proteinase K, 500 µg ml⁻¹ at 56 °C overnight. Day 3: beads were removed in magnetic field. Phenol (pH 4.3) extraction was performed after addition of NaCl (0.2 M final concentration). Ethanol precipitation (in the presence of glycogen): 3 h at -20 °C. The pellet was dissolved in 180 µl H₂O, heated at 75 °C for 3 min and immediately chilled on ice. Samples were treated with DNase I (250 U ml⁻¹) in the presence of RNase inhibitor 300 U ml⁻¹ in ×1 buffer no. 2 (NEB) at 37 °C for 30 min. Phenol (pH 4.3) extraction and ethanol precipitation were repeated. The RNA pellet was dissolved in 50 µl H₂O.

Tobacco acid pyrophosphatase and 5'-phosphate-dependent-exonuclease (Terminator) treatment. Equal amounts of RNA collected from HL-60 cells (as described above) were digested with tobacco acid pyrophosphatase (TAP; Epicentre), Terminator (Epicentre) or no enzyme according to the manufacturer's instructions. RNA was re-extracted in presence of glycogen (Ambion) with acidic phenol (pH 4.3), precipitated with ethanol and re-suspended in ddH₂O.

ecCEBPA, *CEBPA* and 18S expression levels were measured by qRT-PCR using the TaqMan primer sets indicated in Supplementary Methods.

Downregulation of *ecCEBPA*. Three different shRNAs targeting human *ecCEBPA* and scrambled control were designed according to Dharmacon software and cloned into the lentiviral vector pLKO.1 (Sigma-Aldrich), which has a puromycin selection marker. shRNA sequences are shown in Supplementary Methods. Lentiviral particles were produced as described previously⁵⁵. HEK293T cells were co-transfected with either empty vector or the pLKO-shRNA vector and Gag-Pol and Env constructs using Lipofectamine TM 2000 (Invitrogen) according to the manufacturer's recommendation. Virus-containing supernatants were collected 48 and 72 h after transfection and concentrated using a Centricon Plus-70 molecular weight cut-off column (Millipore). Lentiviral transduction was performed in the presence of hexadimethrine bromide (final concentration 8 µg ml⁻¹) in the human myeloid cell line U937. Puromycin (2 µg ml⁻¹) was added to the cultures 2 days after infection. Resistant clones were selected and screened for downregulation of *ecCEBPA* by qRT-PCR.

Upregulation of *ecCEBPA*. The 3' *ecCEBPA* region (R1), the upstream and downstream *ecCEBPA* regions, and the unrelated genomic region (UR, see Supplementary Methods) were cloned into the pBabe retrovirus vector harbouring a puromycin selection marker (Addgene; plasmid 1764). Oligonucleotide sequences used to amplify both regions are shown in Supplementary Methods. K562 cells were transfected with the Amaxa Cell Line Nucleofector Kit V, Program T-003. Puromycin ($2 \mu\text{g ml}^{-1}$) was added to the cultures 2 days after transfection. Resistant clones were selected and screened for upregulation of *ecCEBPA* and the UR by northern blot analysis.

Bisulphite treatment, COBRA and bisulphite sequencing. The methylation profile of the *CEBPA* gene locus was performed by bisulphite sequencing as described previously⁵⁶. In brief, 1 μg of genomic DNA was bisulphite-converted by using the EZ DNA Methylation kit (Zymo Research). Primers and PCR conditions for bisulphite sequencing and COBRA are summarized in Supplementary Methods. For COBRA, PCR products were gel-purified and incubated with BstUI at 60°C for 3 h. The digested DNA was then separated on a 3.5% agarose gel and stained with ethidium bromide. For bisulphite sequencing, PCR products were gel-purified (Qiagen) and cloned into the pGEM-T Easy Vector System (Promega). Sequencing results were analysed using BiQ analyser software⁵⁷. Samples with conversion rate $< 90\%$ and sequences identity $< 70\%$ as well as clonal variants were excluded from our analysis⁵⁵. The minimum number of clones for each sequencing condition was ≥ 6 . Primer sequences are shown in Supplementary Methods.

5-aza-CR treatment. K562 cells were treated with 5-aza-CR⁵⁸ (Sigma-Aldrich) according to the manufacturer's instructions. Medium was refreshed every 48 h. RNA (for RT-PCR) and genomic DNA (for bisulphite sequencing) were isolated after 7 days of treatment.

MassARRAY. Quantitative DNA methylation analysis using the MassARRAY technique was performed by Sequenom as described previously⁴³. In brief, 1 μg of genomic DNA was converted with sodium bisulphite using the EZ DNA methylation kit (Zymo Research), PCR amplified, *in vitro* transcribed and then cleaved by RNase A. The samples were then quantitatively tested for their DNA methylation status using matrix-assisted laser desorption/ionization-time of flight mass spectrometry. The samples were desalted and spotted on a 384-pad SpectroCHIP (Sequenom) using a MassARRAY nanodispenser (Samsung), followed by spectral acquisition on a MassARRAY Analyzer Compact MALDI-TOF MS (Sequenom). The resulting methylation calls were obtained using the EpiTyper software v1.0 (Sequenom) to generate quantitative results for each CpG site or an aggregate of multiple CpG sites. The methylation levels of aggregated multiple CpGs were calculated as the mean of each CpGs methylation value and presented as a percentage. Primer sequences are provided in Supplementary Data 1.

Transfection of human DNMT1-haemagglutinin tag construct and western blot analysis. The human DNMT1-haemagglutinin-tagged cloned into the expression vector pCDNA3.1 was a kind gift from S. Baylin. HEK293T cells were transfected using Lipofectamine 2000, and 2 days later collected for western blot analysis. Single-cell suspensions were lysed with modified radioimmunoprecipitation assay buffer, and whole-cell lysates separated on 6% SDS-PAGE gels. Immunoblots were stained with the following primary antibodies: DNMT1 (1:5,000, Abcam) and HSP90 (1:2,000, BD Biosciences). All secondary antibodies were horseradish peroxidase (HRP)-conjugated (Santa Cruz) and diluted 1:5,000 for rabbit-HRP, and 1:3,000 for mouse-HRP. Western blot analysis for the HCT116 hypomorphic and wild-type cell lines were similarly performed.

T7 polymerase-induced transcription. The T7 expression system is based on technology developed at Brookhaven National Laboratory under contract with the US Department of Energy and is the subject of patents and pending applications. Full information may be obtained from the Office of Intellectual Property and Sponsored Research, Brookhaven National Laboratory. Maps of T7 polymerase constructs are presented in ref. 59. In brief, the murine RAW 264.7 cell line was stably transfected with a construct carrying the human genomic segment under T7 promoter control (derived from pBlueScript plasmid; Agilent). After selection in G418, individual clones were transfected with T7 polymerase-expressing mammalian constructs and were tested by COBRA for genomic methylation.

EMSAs and K_d determination. DNA and RNA oligonucleotides (15 pmol) were end-labelled with [γ - ^{32}P]ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs). Reactions were incubated at 37°C for 1 h and then passed through G-25 spin columns (GE Healthcare) according to the manufacturer's instructions to remove unincorporated radioactivity. Labelled samples were gel-purified on 10% polyacrylamide gels. Binding reactions were carried out in 10- μl volumes in the following buffer: 5 mM Tris, pH 7.4, 5 mM MgCl_2 , 1 mM dithiothreitol (DTT), 3% (v/v) glycerol, 100 mM NaCl. Various amounts (0.021–0.156 μM) of purified DNMT1 protein (BPS Bioscience) were incubated with 1.1 nM of ^{32}P -labelled double-stranded DNA and single/double-stranded RNAs. In the competitive assay, a fixed amount of protein and increasing amounts of competitors (double-stranded DNA or poly(dI-dC)) were used. All reactions were

assembled on ice and incubated at room temperature. Samples were separated on 6% native polyacrylamide gels ($0.5 \times \text{Tris/Borate/EDTA}$ (TBE); 4°C ; $\sim 3 \text{ h}$ at 140 V). Gels were dried and exposed to X-ray film and/or PhosphorImager screens. Quantification was done with ImageQuant software. For affinity assays, the per cent shifted species was determined as follows: the migration of the labelled DNA in this reaction was defined as zero per cent shifted and the ratio of the PhosphorImager counts in the area of the lane above this band to the total counts in the lane was defined as background and subtracted from all other lanes. This band represented total input. Subsequent lanes containing DNMT1–nucleic acid complexes were treated identically, and the percentage complex formation was calculated as follows: $[\% \text{ bound complex}] = (1 - ((\text{unbound} - \text{background})/(\text{input} - \text{background})))$. All experiments contained a control reaction lacking DNMT1. The percentage complex formation was plotted as a function of DNMT1 concentration using nonlinear regression analysis performed with Prism 4.0a software. RNA and DNA oligonucleotides used in EMSA are shown in Fig. 3a and Supplementary Fig. 3d and listed in Supplementary Methods.

***ecCEBPA* binding to GST–DNMT1 fragments.** GST and GST–DNMT1 fragments were expressed and purified by glutathione sepharose affinity beads (GE Healthcare Life Sciences) as described previously⁶⁰. Protein concentrations for the recombinant GST and GST–DNMT1 fragments were determined by gel electrophoresis and subsequent Coomassie staining and densitometry. ^{32}P end-labelling of *ecCEBPA* oligonucleotides was carried out at 37°C for 30 min in a total volume of 50 μl . The reaction contained 50 pmol *ecCEBPA*, adenosine 5'-triphosphate, [γ - ^{32}P] (specific activity 3,000 Ci mmol^{-1} , Perkin Elmer) and 20 U of T4 kinase (New England Biolabs) mixed in assay buffer (70 mM Tris-HCl, 10 mM MgCl_2 , 5 mM DTT, pH 7.6). The labelled *ecCEBPA*- ^{32}P was purified with illustraMicroSpin G-25 Columns (GE Healthcare Life Sciences) according to manufacturer's specifications. Equal amounts of the GST and GST–DNMT1 fragments were mixed with 5 μl *ecCEBPA*- ^{32}P , in duplicate, and incubated at 37°C for 10 min in a total volume of 25 μl of reaction buffer (50 mM Tris-HCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, pH 7.8). The sepharose beads were then washed four times in phosphate-buffered solution and placed in 3 ml of scintillation fluid and bound ^{32}P was measured for 1 min. All measurements were normalized to ^{32}P readings for the corresponding input ^{32}P -*ecCEBPA*.

***In vitro* transcription–methylation assay.** *In vitro* transcription–methylation assays were performed on hmDNA (described in Supplementary Information, legend to Supplementary Fig. 4) in the presence or absence of 5 U of human DNMT1 enzyme (New England Biolabs) and 5 U of T7 RNA polymerase (Promega) or 5 U of *E. coli* RNA polymerase sigma-saturated holoenzyme (Epicentre). Reactions were performed in DNMT1 buffer according to the manufacturer's recommendations supplemented with rNTPs (ribonucleotide triphosphates) and 1.25 mM MgCl_2 , including the 'DNMT1 only' reaction. This predetermined concentration of Mg^{2+} cations is high enough to sustain activity of RNA polymerases and low enough not to inhibit DNMT1 activity.

DNMT1 *in vitro* methylation assay. DNMT1 enzymatic assays were carried out in duplicate at 37°C for 30 min in a total volume of 25 μl . The reaction contained S-adenosyl-L-[methyl- ^3H]methionine (AdoMet) (specific activity 18 Ci mmol^{-1} , Perkin Elmer), 200 ng of substrate DNA, recombinant DNMT1 enzyme (2.5 pmol) and *ecCEBPA* (2.5 pmol) mixed in assay buffer (50 mM Tris-HCl, 1 mM DTT, 1 mM EDTA, 5% glycerol, pH 7.8). Methyltransferase reactions were snap-frozen in an ethanol-dry ice bath. The entire reaction volume (25 μl) was spotted on 2.5-cm DE81 membranes (GE Healthcare). These membranes were processed by washes in $3 \times 1 \text{ ml}$ of 0.2 M ammonium bicarbonate, $3 \times 1 \text{ ml}$ water and $3 \times 1 \text{ ml}$ ethanol. Processed membranes were air-dried, placed in 3 ml of scintillation fluid and tritium incorporation was measured for 1 min. Background subtraction (no DNA substrate) was performed for all experimental sample counts.

RIP-seq. Total RNA immunoprecipitated with DNMT1 antibody (Abcam) or IgG (Sigma-Aldrich) was processed for sequencing as described in ref. 61 with some modifications. Double-stranded cDNA was synthesized using the Just cDNA Double-Stranded cDNA Synthesis Kit (Agilent Technology) according to the manufacturer's instructions. Illumina sequencing libraries were constructed from these cDNA using a ChIP-seq sample preparation kit (Illumina) with minor modifications. Illumina paired-end adaptor and PCR primers were used to replace the single-read adaptor and primers in the kit. Constructed libraries were subjected to a final size-selection step on 10% Novex TBE gels (Invitrogen). DNA fragments of 175–200 base pairs (bp) were excised from a SYBR-green-stained gel. DNA was recovered from the gel and quantified following Illumina's qPCR quantification protocol. Paired-end sequencing of these libraries was then performed on an Illumina GA IIx to achieve 2×76 -bp reads. Paired-end reads were trimmed to 50 bp and aligned to the reference genome hg19 using BWA⁶² with the following parameters: `bwa aln -o 1 -l 25 -k 2; bwa sampe -o 200`. To estimate a normalization factor (alpha) between the immunoprecipitations, the genome was divided into course bins (10 kb) and reads were counted for DNMT1 RIP and IgG control in

each bin. A linear regression was fitted across all non-zero bins and the slope of the regression was used as a scaling factor (α) to normalize the RIP and control libraries. To identify distinct regions specifically bound by DNMT1, all downstream analyses were conducted on a set of regions derived by aggregating overlapping DNMT1 RIP reads into contiguous intervals. Each DNMT1 interval was tested for significance by comparing the number of reads within the interval with the number of reads in the same region of the IgG control, multiplied by the previously estimated scaling factor, α (exact binomial test, $P = 0.5$). Multiple tests were corrected by Benjamini–Hochberg. In total, 16,186 intervals (representing the start and end boundaries of contiguous, overlapping reads) were determined to be significantly enriched in the DNMT1 RIP as compared to the IgG control ($P < 0.0001$; $q < 0.0001$). A false discovery rate of 7.5% was determined by determining the number of significantly enriched intervals in the IgG immunoprecipitate using DNMT1 as a control. Significantly enriched DNMT1 intervals have a mean length of 347 bp and a median of 67 reads per interval. Every peak represents an interval with a ‘height’ value: the sum of all reads within an interval. All peaks were annotated with CEAS³³ build on RefSeq hg19. All DNMT1 RIP-seq peaks were also annotated using the HOMER pipeline (version 4.2)³⁴ which provide a comprehensive RNAs database (coding and non-coding, including miRNA, small nucleolar RNA, ribosomal RNA, small nuclear RNA, transfer RNA, etc.).

A peak was considered as belonging to a gene if located in the gene body or 3 kb up- or downstream the gene (gene loci). Altogether, 6,042 gene loci were covered by at least one significant RIP-seq peak.

RNA-seq. Total RNA was extracted with TRI Reagent (MRC). RNA samples were treated with 10 U of DNase I (Roche Applied Science) per 3 μ g of total RNA at 37 °C for 1 h in the presence of RNaseA inhibitor. Non-polyadenylated RNA fractions were selected with the MicroPoly(A) Purist purification kit (Ambion). Total and non-polyadenylated RNA were depleted of rRNA with Ribo-Zero™ Magnetic Gold Kit (Epicentre). Double-stranded cDNA libraries were constructed using ScriptSeq v2 RNA-Seq Library Preparation Kit (Epicentre) followed by duplex specific nuclease (DSN) normalization (Evrogen). DSN-treated libraries were subjected to final size selection in 3% agarose gel. 250–500-bp fragments were excised and recovered using the Qiaquick Gel Extraction Kit (Qiagen). Libraries were sequenced (1 per lane) on a Hi-Seq-2000 Illumina instrument. Raw read sequences were deposited in GEO (accession number GSE32153). 2.96×10^8 reads from the HL-60 and 3.75×10^8 76-bp paired-end reads from the HL-60 poly(A)⁺-depleted RNA were aligned to the human genome hg19 (UCSC release) using Tophat2 (ref. 63). Aligned reads were assembled into individual full-length transcripts using Cufflinks v2.0.2 (ref. 63) and a merged assembly was created from the two assemblies and additionally, all level-1 and -2 transcripts from the Gencode v11 catalogue⁶⁴ using Cuffmerge⁶⁶. To confirm the transcription of the significant DNMT1 RIP-seq peaks, we overlapped the peak intervals with the RIP-seq assemblies using the BEDtools⁶⁰ intersect BED utility.

RRBS. RRBS was performed as described⁴⁴. In brief, high-quality genomic DNA was isolated from the myeloid cell line HL-60. DNA was digested with MspI (NEB), a methylation-insensitive enzyme that cuts C⁺CGG. Digested DNA was size selected on a 4% NuSieve 3:1 agarose gel (Lonza). For each sample, two slices containing DNA fragments of 40–120 bp and 120–220 bp, respectively, were excised from the unstained preparative portion of the gel. These two size fractions were kept apart throughout the procedure and mixed 1:2 for the final sequencing. Pre-annealed Illumina adaptors containing 5'-methylcytosine instead of cytosine were ligated to size-selected MspI fragments. Adaptor-ligated fragments were bisulphite-treated using the EZ DNA Methylation kit (Zymo Research). The products were PCR amplified, size selected and sequenced on the Illumina GAIIX at a reading length of 36 bp. Sequencing reads were mapped to the reference genome hg19 using RRBSmap⁴⁵ allowing two mismatches. Reads from replicates were merged and processed as described previously⁴⁶. We considered only CpG located in regions with a depth of coverage greater than three reads. The β -score of CpG methylation in a given position is the ratio of methylated CpGs within the total number of CpGs through all reads. The level of gene methylation is the mean of all CpG β -scores within –2 kb from the TSS to the end of first intron; for intronless genes, the entire gene body was considered. Genes with less than three sequenced CpGs in the promoter or less than three sequenced CpGs in the first exon–intron were excluded.

For RRBS in R1- and the UR-overexpressing cells bisulphite sequenced UR–R1 genomes were binned at 100-bp intervals using the R-Bioconductor/methylKit ‘tileMethylCounts’ function (<http://code.google.com/p/methylkit/>)⁴⁷. The level of differential methylation was computed by comparing all sequenced CpG sites within the overlapping bins between the two samples (R1 and UR). The significant differentially methylated bins were obtained using the Fisher’s test from the R-Bioconductor/methylKit ‘calculateDiffMeth’ function (q value < 0.01 and methylation difference $\geq 50\%$). A gene was considered differentially methylated

if the region including the promoter (–2 kb from the TSS) and first exon was overlapped by at least one significant differentially methylated bin. In total, 11,844 promoter/first exon regions were analysed.

RNA expression profiling. RNA isolated from HL-60 cells was used for sample amplification and labelling using the Whole Transcriptome assay reagent kits from Affymetrix. 10 μ g of labelled RNA was hybridized on Affymetrix GeneChip Human Gene 1.0 ST array. Hybridization, washing, staining and scanning were carried out as recommended by the manufacturer. Each hybridization was performed in triplicate. Washes and staining were performed through the Fluidics Station 400 and the GeneChip Scanner 3000 (Affymetrix) was used to measure the fluorescence intensity emitted by the labelled target. Raw data processing was performed using the Affymetrix GeneChip Operating Software (GCOS). Microarrays were RMA normalized using ‘affy’, an R-Bioconductor library⁶⁵. *CEBPA* expression was used as a threshold to define expressing (\log_2 score above 4) and non-expressing (\log_2 score below 4) genes for further analysis.

GO and pyknons comparison. GO analysis was performed with DAVID⁶⁶. We focused our analysis on biological process annotations. GO enrichment was scored using the Benjamini–Hochberg-corrected P value. DNMT1 RIP-seq-specific peaks were compared to the human pyknons database released in January 2013 (https://cm.jefferson.edu/tools_and_downloads/pyknons.html).

Data integration. We used the Ref-seq transcripts database built on hg19 (UCSC release) as a genome annotation reference for Rip-seq, RRBS and microarray expression experiments. We selected only the longest transcripts. Accordingly, the number of 40,857 RefSeq IDs was reduced to 23,250 transcript IDs. Then, we annotated all RIP-seq peaks against the gene loci, which includes exonic, intronic and UTR regions plus 3 kb upstream of the TSS and 3 kb downstream of the transcription end site regions. We identified 6,042 gene loci with DNMT1 RIP-seq peaks and 17,208 gene loci without DNMT1 RIP-seq peaks. Finally, we focused our study on gene loci covered by the RRBS. We identified 4,833 gene loci with DNMT1 RIP-seq peaks and covered by RRBS and 10,973 gene loci without DNMT1 RIP-seq peaks and covered by RRBS. We plotted genes within each group against expression and methylation profile. Using *CEBPA* levels of expression as a cut-off threshold, we defined genes as ‘expressed’ or ‘low or not expressed’ if the \log_2 score was above or below 4, respectively, and as ‘hypomethylated’ and ‘methylated’ genes with mean of all CpG scores below and above 50%, respectively. We identified by this approach four clusters in each group. In DNMT1-unbound group clusters: A (expressed, hypomethylated genes; 23.04%), B (low or not expressed, methylated genes; 51.45%), E (expressed, methylated genes; 10.45%) and F (low or not expressed, hypomethylated genes; 15.06%). In DNMT1-bound group clusters: C (expressed, hypomethylated genes; 56.64%), D (low or not expressed, methylated genes; 12.04%), G (expressed, methylated genes; 12.08%), H (low or not expressed, hypomethylated genes; 19.24%). In all RRBS-covered group clusters: I (expressed, hypomethylated genes; 33.32%), J (low or not expressed, methylated genes; 39.40%), K (expressed, methylated genes; 10.95%) and L (not or low expressed, hypomethylated genes; 16.34%).

Statistical analysis. Methylation changes of clones analysed by bisulphite sequencing were calculated using the Fisher’s exact test (GraphPad Prism Software). Methylation changes assessed by MassARRAY were calculated using a student’s t -test (GraphPad Prism Software). The statistical evaluation of DNMT1–RNA interaction versus expression and methylation was estimated using the student’s t -test (box-plots; Supplementary Fig. 5d). The overrepresentation of genes in clusters B and C following our hypothesis against those which did not, was assessed using a 2-sample proportion test (Fig. 5b). P values for t -test and 2-sample proportion test—‘ t .test’ and ‘prop.test’, respectively—were calculated by the R functions (<http://www.r-project.org>). Values of $P \leq 0.05$ were considered statistically significant. The mean \pm s.d. of two or more replicates is reported.

49. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1982).
50. Blobel, G. & Potter, V. R. Nuclei from rat liver: isolation method that combines purity with high yield. *Science* **154**, 1662–1665 (1966).
51. Sambrook, J. & Russell, D. W. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
52. Mitchell, J. A. & Fraser, P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* **22**, 20–25 (2008).
53. Dieci, G., Fiorino, G., Castelnovo, M., Teichmann, M. & Pagano, A. The expanding RNA polymerase III transcriptome. *Trends Genet.* **23**, 614–622 (2007).
54. Pagano, A. *et al.* New small nuclear RNA gene-like transcriptional units as sources of regulatory transcripts. *PLoS Genet.* **3**, e1 (2007).
55. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493–501 (2003).
56. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89**, 1827–1831 (1992).
57. Bock, C. *et al.* BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* **21**, 4067–4068 (2005).

58. Christman, J. K. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483–5495 (2002).
59. Dunn, J. J., Kripl, B., Bernstein, K. E., Westphal, H. & Studier, F. W. Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene* **68**, 259–266 (1988).
60. Estève, P. O. *et al.* Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* **20**, 3089–3103 (2006).
61. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
63. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
64. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (suppl. 1), S4.1–9 (2006).
65. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
66. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

An Earth-sized planet with an Earth-like density

Francesco Pepe¹, Andrew Collier Cameron², David W. Latham³, Emilio Molinari^{4,5}, Stéphane Udry¹, Aldo S. Bonomo⁶, Lars A. Buchhave^{3,7}, David Charbonneau³, Rosario Cosentino^{4,8}, Courtney D. Dressing³, Xavier Dumusque³, Pedro Figueira⁹, Aldo F. M. Fiorenzano⁴, Sara Gettel³, Avet Harutyunyan⁴, Raphaëlle D. Haywood², Keith Horne², Mercedes Lopez-Morales³, Christophe Lovis¹, Luca Malavolta^{10,11}, Michel Mayor¹, Giusi Micela¹², Fatemeh Motalebi¹, Valerio Nascimbeni¹¹, David Phillips³, Giampaolo Piotto^{10,11}, Don Pollacco¹³, Didier Queloz^{1,14}, Ken Rice¹⁵, Dimitar Sasselov³, Damien Ségransan¹, Alessandro Sozzetti⁶, Andrew Szentgyorgyi³ & Christopher A. Watson¹⁶

Recent analyses^{1–4} of data from the NASA Kepler spacecraft⁵ have established that planets with radii within 25 per cent of the Earth's (R_{\oplus}) are commonplace throughout the Galaxy, orbiting at least 16.5 per cent of Sun-like stars¹. Because these studies were sensitive to the sizes of the planets but not their masses, the question remains whether these Earth-sized planets are indeed similar to the Earth in bulk composition. The smallest planets for which masses have been accurately determined^{6,7} are Kepler-10b ($1.42R_{\oplus}$) and Kepler-36b ($1.49R_{\oplus}$), which are both significantly larger than the Earth. Recently, the planet Kepler-78b was discovered⁸ and found to have a radius of only $1.16R_{\oplus}$. Here we report that the mass of this planet is 1.86 Earth masses. The resulting mean density of the planet is 5.57 g cm^{-3} , which is similar to that of the Earth and implies a composition of iron and rock.

Every 8.5 h, the star Kepler-78 (first known as TYC 3147-188-1 and later designated KIC 8435766) presents to the Earth a shallow eclipse consistent⁸ with the passage of an orbiting planet with a radius of $1.16 \pm 0.19R_{\oplus}$. A previous study⁸ demonstrated that it was very unlikely that these eclipses were the result of a massive companion either to Kepler-78 itself or to a fainter star near its position on the sky. Judging from the absence of ellipsoidal light variations⁸ of the star, the upper limit on the mass of the planet is 8 Earth masses (M_{\oplus}). In addition to its diminutive size, the planet Kepler-78b is interesting because the light curve recorded by the Kepler spacecraft reveals the secondary eclipse of the planet behind the star as well as the variations in the light received from the planet as it orbits the star and presents different hemispheres to the observer. These data enabled constraints⁸ to be put on the albedo and temperature of the planet. A direct measurement of the mass of Kepler-78b would permit an evaluation of its mean density and, by inference, its composition, and motivated this study.

The newly commissioned HARPS-N⁹ spectrograph is the Northern Hemisphere copy of the HARPS¹⁰ instrument, and, like HARPS, HARPS-N allows scientific observations to be made alongside thorium–argon emission spectra for wavelength calibration¹¹. HARPS-N is installed at the 3.57-m Telescopio Nazionale Galileo at the Roque de los Muchachos Observatory, La Palma Island, Spain. The high optical efficiency of the instrument enables a radial-velocity precision of 1.2 m s^{-1} to be achieved in a 1-h exposure on a slowly rotating late-G-type or K-type dwarf star with stellar visible magnitude $m_v = 12$. By observing standard stars of known radial velocity during the first year of operation of HARPS-N, we estimated it to have a precision of at least 1 m s^{-1} , a value which is roughly half the semi-amplitude of the signal expected for Kepler-78b should the planet have a rocky composition. We began an intensive

observing campaign (Methods) of Kepler-78 ($m_v = 11.72$) in May 2013, acquiring HARPS-N spectra of 30-min exposure time and an average signal-to-noise ratio of 45 per extracted pixel at 550 nm (wavelength bin of 0.00145 nm). From these high-quality spectra, we estimated^{12,13} the stellar parameters of Kepler-78 (Methods and Extended Data Table 1). Our estimate of the stellar radius, $R_* = 0.737^{+0.034}_{-0.042} R_{\odot}$, is more accurate than any previously known⁸ and allows us to refine the estimate of the planetary radius.

In the Supplementary Data, we provide a table of the radial velocities, the Julian dates, the measurement errors, the line bisector of the cross-correlation function, and the Ca II H-line and K-line activity indicator¹⁴, $\log(R'_{\text{HK}})$. The radial velocities (Fig. 1) show a scatter of 4.08 m s^{-1} and a peak-to-trough variation of 22 m s^{-1} , which exceeds the estimated average internal (photon-noise) precision, of 2.3 m s^{-1} . The excess scatter is probably due to star-induced effects including spots and changes in the convective blueshift associated with variations in the stellar activity. These effects may cause an apparent signal at intervals corresponding to the stellar rotational period and its first and second harmonics. To separate this signal from that caused by the planet, we proceeded to estimate the rotation period of the star from the de-trended light curve from Kepler (Methods). Our estimate, of

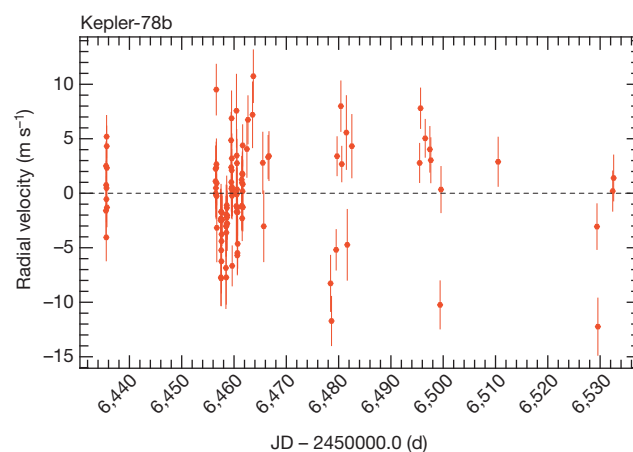


Figure 1 | Radial velocities of Kepler-78 as a function of time. The error bars indicate the estimated internal error (mainly photon-noise-induced error), which was $\sim 2.3 \text{ m s}^{-1}$ on average. The signal is dominated by stellar effects. The raw radial-velocity dispersion is 4.08 m s^{-1} , which is about twice the photon noise. JD, Julian date.

¹Observatoire Astronomique de l'Université de Genève, 51 chemin des Maillettes, 1290 Versoix, Switzerland. ²Scottish Universities Physics Alliance, School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK. ³Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ⁴INAF - Fundación Galileo Galilei, Rambla José Ana Fernández Pérez 7, 38712 Breña Baja, Spain. ⁵INAF - IASF Milano, via Bassini 15, 20133 Milano, Italy. ⁶INAF - Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese, Italy. ⁷Centre for Star and Planet Formation, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen, Denmark. ⁸INAF - Osservatorio Astrofisico di Catania, via Santa Sofia 78, 95125 Catania, Italy. ⁹Centro de Astrofísica, Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal. ¹⁰Dipartimento di Fisica e Astronomia "Galileo Galilei", Università di Padova, Vicolo dell'Osservatorio 3, 35122 Padova, Italy. ¹¹INAF - Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy. ¹²INAF - Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, 90124 Palermo, Italy. ¹³Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. ¹⁴Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. ¹⁵Scottish Universities Physics Alliance, Institute for Astronomy, Royal Observatory, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK. ¹⁶Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK.

~ 12.6 d, is consistent with a previous estimate⁸. The power spectral density of the de-trended light curve also shows strong harmonics at respective periods of 6.3 and 4.2 d. We note that these timescales are much longer than the orbital period of the planet.

To estimate the radial-velocity semi-amplitude, K_p , due to the planet, we proceeded under the assumption that K_p was much larger than the change in radial velocity arising from stellar activity during a single night. This is a reasonable assumption, because a typical 6-h observing sequence spans only 2% of the stellar rotation cycle. Furthermore, the relative phase between the stellar signal and the planetary signal changes continuously, and so we expect that the contribution from stellar activity should average out over the three-month observing period. Assuming a circular orbit, we modelled the radial velocity of the i th observation (gathered at time t_i on night d) as $v_{i,d} = v_{0,d} - K_p \sin(2\pi(t_i - t_0)/P)$, where t_0 is the epoch of mid-transit and P is the orbital period, each held fixed at the values derived from the photometry, and $v_{0,d}$ is the night- d zero point (an offset value estimated independently for each night). We solved for the values of K_p and $v_{0,d}$ by minimizing the χ^2 function (least-squares minimization), assuming white noise and weighting the data according to the

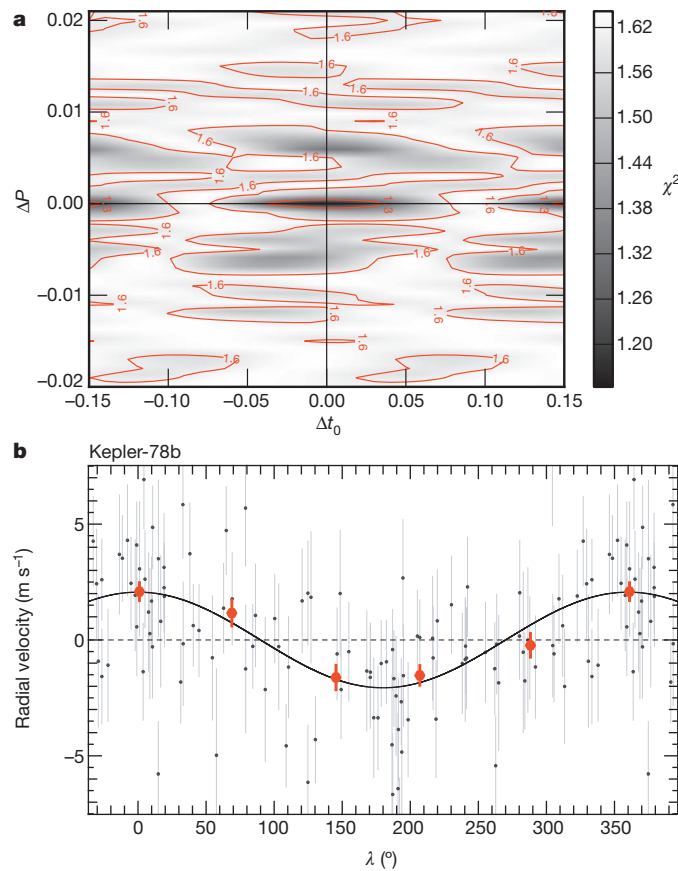


Figure 2 | Model fit of the radial velocities of Kepler-78. **a**, Greyscale contour plot of the χ^2 surface around the ephemeris for the period and mid-time of transit from ref. 8 (origin of plot). ΔP and Δt_0 represent the departure from the expected ephemeris in units of days. The position of the minimum of the residuals perfectly matches the expected values. **b**, Phase-folded radial velocities and fitted Keplerian orbit of the signal induced by Kepler-78b after removal of the modelled stellar noise components. λ , mean longitude. The transit occurs at $\lambda_T = 90^\circ$. We note that the higher number of data points at maximum and minimum radial velocity is a direct consequence of our strategy of observing at quadrature, where the information on amplitude is highest. The red dots show the radial velocities and the corresponding errors when binned over the orbital phase. The error bars of the individual measurements indicate the estimated internal errors (photon-noise-induced error), of ~ 2.3 m s⁻¹ on average. The error bars on the averaged data points essentially represent the internal error of an individual measurement divided by the square root of the number of averaged measurements.

inverse variances derived from the HARPS-N noise model. This is a technique similar to the one used in a recent study¹⁵ of another low-mass transiting planet, CoRoT-7b.

With this technique, we measure a preliminary value of $K_p = 2.08 \pm 0.32$ m s⁻¹, implying a detection significance of 6.5σ . We confirmed that the radial-velocity signal is consistent with the photometric transit ephemeris by repeating the χ^2 optimization over a grid of orbital frequencies and times of mid-transit (Fig. 2a). This confirms that the values of P and t_0 from the Kepler light curves coincide with the lowest χ^2 minimum in the resulting period–phase diagram.

The Kepler light curve evolves on a timescale of weeks. We therefore expect the stellar-activity-induced radial-velocity signals to remain coherent on the same timescale. To explore this, we fitted the radial velocities with a sum of Keplerian models at different periods, one of which we expect to correspond to the planetary orbit and the others (which we left as free parameters) to represent the effects of stellar activity. Using the Bayesian information criterion as our discriminant (Methods), we found that a model with three Keplerian functions was sufficient to explain the data. The 4.2-d period of the second Keplerian function corresponds to the second harmonic of the photometrically determined stellar rotation period. The period of 10 d for the third Keplerian also appears as a prominent peak in the generalized Lomb–Scargle analysis¹⁶ of the radial-velocity data. Strong peaks at similar periods are also present in periodograms of the Ca II HK activity indicator, the line bisector and the full-width at half-maximum of the cross-correlation function¹¹ (Extended Data Fig. 1). We conclude that both the 4.2-d and 10-d signals probably have stellar causes.

The best three-Keplerian fit of the data yields an estimate of $K_p = 1.96 \pm 0.32$ m s⁻¹ and residuals with a dispersion of 2.34 m s⁻¹, very close to the internal noise estimates. In Fig. 2b, we show the phase-folded radial velocities after removal of the stellar components, plotted along with the best-fit Keplerian at the planetary orbital period. The orbital parameters are given in Table 1. Having settled on the three-component model, we estimated the planetary mass and density by conducting a Markov chain Monte Carlo (MCMC) analysis (Methods). In this analysis, we adopted previously published⁸ values of P and t_0 as Gaussian priors. We replaced the planet radius, R_p , with the published estimate⁸ of the ratio $k = R_p/R_*$ and our determination of R_* . The planetary radius then becomes an output of the MCMC analysis. By adopting the mode of the distributions, we find a planet mass of $m_p = 1.86^{+0.38}_{-0.25} M_\oplus$, a radius of $R_p = 1.173^{+0.159}_{-0.089} R_\oplus$ and a planet density of $\rho_p = 5.57^{+3.02}_{-0.31}$ g cm⁻³. These estimates include the contribution from

Table 1 | Planetary system parameters of Kepler-78 b

Planetary system parameter	Kepler-78b
Reference epoch, T_0 (BJD _{TTC})	2,456,465.076392
Orbital inclination, i (°)*	79^{+9}_{-14}
Systemic radial velocity, γ (km s ⁻¹)†	-3.5084 ± 0.0008
Orbital period, P (d)†	0.3550 ± 0.0004
Mean longitude, λ_0 , at epoch T_0 (°)†	293 ± 13
Eccentricity, e †	0
Radial-velocity semi-amplitude, K_p (m s ⁻¹)†	1.96 ± 0.32
Planetary mass, m_p (M_\oplus)‡	$1.86^{+0.38}_{-0.25}$
Planetary radius, R_p (R_\oplus)‡	$1.173^{+0.159}_{-0.089}$
Planetary mean density, ρ_p (g cm ⁻³)‡	$5.57^{+3.02}_{-0.31}$
Semi-major axis, a (AU)†	0.0089
Surface temperature, T (K)*	1500–3000
Number of measurements, N_{meas}	109
Time span of observations (d)	97.1
Radial-velocity dispersion ($O - C$) (m s ⁻¹)†	2.34
Reduced χ^2 †	1.12 ± 0.07

The parameters are determined from radial velocities. BJD_{TTC} indicates barycentric Julian date expressed in coordinated universal time. λ_0 denotes the mean longitude at the mean date of the observing campaign (reference epoch, T_0). This coordinate has the advantage of not being degenerate for low eccentricities. Our choice of T_0 reduces correlations between adjusted parameters. $O - C$ is the standard deviation of the difference between the observed and computed (modelled) data. Stated uncertainties represent the 68.3% confidence interval.

* Taken from the discovery paper⁸. † Fitted orbital parameters (maximum likelihood). ‡ Mode and confidence interval of the real distribution deduced from the MCMC analysis.

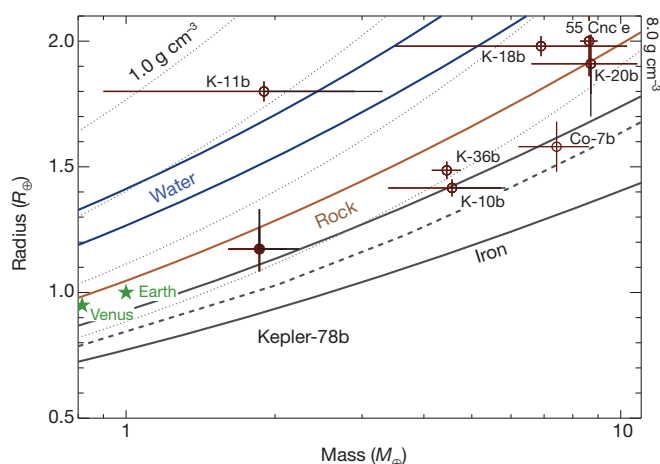


Figure 3 | The hot, rocky planet Kepler-78b placed on a planetary mass–radius diagram. Here masses and radii represent the modes of the corresponding distributions derived from the MCMC analysis, and the error bars represent the 68.3% confidence interval (1σ). For comparison, Earth and Venus are indicated in the same diagram by star symbols. The other exoplanets shown are those for which mass and radius have been estimated (K, Kepler; Co, CoRoT; 55 Cnc, 55 Cancri). From top to bottom, the solid lines show mass and radius for planets consisting of pure water, 50% water, 100% silicates, 50% silicates and 50% iron core, and 100% iron, as computed with the theoretical models of ref. 18. The dashed line shows the maximum mantle-stripping models from ref. 21; that is, Earth-like exoplanets made of pure iron are not expected to form around normal stars. From top to bottom, the dotted lines represent mean densities of 1, 2, 4 and 8 g cm^{-3} .

the uncertainty in the stellar mass. The uncertainty in the density is dominated by the uncertainty in k . Our values for m_p and ρ_p are consistent with those from an independent study¹⁷.

In terms of mass, radius and mean density, Kepler-78b is the most similar to the Earth among the exoplanets for which these quantities have been determined. We plot the mass–radius diagram in Fig. 3. By comparing our estimates of Kepler-78b with theoretical models¹⁸ of internal composition, we find that the planet has a rocky interior and most probably a relatively large iron core (perhaps comprising 40% of the planet by mass). We note that in the part of the mass–radius diagram where Kepler-78b lies, there is a general agreement between models and little or no degeneracy. The extreme proximity of the planet to its star, resulting in a high surface temperature and ultraviolet irradiation, would preclude there being a low-molecular-weight atmosphere: any water or volatile envelope that Kepler-78b might have had at formation should have rapidly evaporated¹⁹. Kepler-78b is also similar to larger high-density, hot exoplanets (Kepler-10b (ref. 6), Kepler-36b (ref. 7) and CoRoT-7b (ref. 20)), in that in the mass–radius diagram it is not below the lower envelope of mantle-stripping models²¹ that tend to enhance the fraction of the planet’s iron core. At present, Kepler-78b is the extrasolar planet whose mass, radius and likely composition are most similar to those of the Earth. However, it differs from the Earth notably in its very short orbital period and correspondingly high temperature.

The observations of Kepler-78 have shown the potential of the much-anticipated HARPS-N spectrograph. It will have a crucial role in the characterization of the many Kepler planet candidates with radii similar to that of the Earth. By acquiring and analysing a large number of precise radial-velocity measurements, we can learn whether Earth-sized planets (typically) have Earth-like densities (and, by inference, Earth-like compositions), or whether even small planets have a wide range of compositions, as has recently been established^{22,23} for their larger kin.

METHODS SUMMARY

In the case of Kepler-78, the planet-induced radial-velocity variation is small compared with the stellar jitter. If their periodicities are very different, however, it is easy to de-correlate the signals and determine the radial-velocity amplitude of the

planet. We used the Kepler light curve of Kepler-78 to measure the stellar rotational period. After de-trending^{24,25} the photometry, we computed its power spectral density, which immediately revealed excess power at a period of 12.6 d, the rotational period of the star.

The HARPS-N observations of Kepler-78 yielded not only radial velocities but also high-resolution spectra, which we combined into a spectrum with a high signal-to-noise ratio. By applying the stellar parameter classification pipeline¹², we derived precise stellar parameters. In particular, we re-determined the stellar radius to be $R_* = 0.737^{+0.034}_{-0.042} R_\odot$, with smaller uncertainties than in the value in the discovery paper⁸.

For the purpose of measuring the signal induced by Kepler-78b, several models can be applied to the data, which may all lead to similar results. However, not all of the models represent the data with the same quality. We therefore used the Bayesian information criterion^{26–28} to determine which model matches the data best. This analysis led us to select a three-Keplerian model with two sinusoids (zero-eccentricity Keplerians) for the planet and the 4.2-d stellar signal, respectively, and one Keplerian with non-zero eccentricity for the stellar signal at 10 d.

Once a model has been selected, it is adjusted by a least-squares fit to the data. This approach leads to the maximum-likelihood solution but does not provide all statistically relevant solutions and the distributions of their parameters. We used an MCMC analysis to determine the distribution of all orbital and planetary parameters, in particular the planetary mass and density, and to determine their respective errors.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 September; accepted 14 October 2013.

Published online 30 October 2013.

1. Fressin, F. *et al.* The false positive rate of Kepler and the occurrence of planets. *Astrophys. J.* **766**, 81–100 (2013).
2. Petigura, E. A., Marcy, G. W. & Howard, A. W. A plateau in the planet population below twice the size of Earth. *Astrophys. J.* **770**, 69–89 (2013).
3. Swift, J. J. *et al.* Characterizing the cool KOIs. IV. Kepler-32 as a prototype for the formation of compact planetary systems throughout the galaxy. *Astrophys. J.* **764**, 105–118 (2013).
4. Dressing, C. D. & Charbonneau, D. The occurrence rate of small planets around small stars. *Astrophys. J.* **767**, 95–114 (2013).
5. Batalha, N. M. *et al.* Planetary candidates observed by Kepler. III. Analysis of the first 16 months of data. *Astrophys. J.* **204** (Suppl.), 24–44 (2011).
6. Batalha, N. M. *et al.* Kepler’s first rocky planet: Kepler-10b. *Astrophys. J.* **729**, 27–47 (2011).
7. Carter, J. A. *et al.* Kepler-36: a pair of planets with neighboring orbits and dissimilar densities. *Science* **337**, 556–559 (2012).
8. Sanchis-Ojeda, R. *et al.* Transits and occultations of an Earth-sized planet in an 8.5-hour orbit. *Astrophys. J.* **774**, 54–62 (2013).
9. Cosentino, R. *et al.* Harps-N: the new planet hunter at TNG. *Proc. SPIE* **8446**, 84461V (2012).
10. Mayor, M. *et al.* Setting new standards with HARPS. *Messenger* **114**, 20–24 (2003).
11. Baranne, A. *et al.* ELODIE: a spectrograph for accurate radial velocity measurements. *Astron. Astrophys.* **119** (Suppl.), 373–390 (1996).
12. Buchhave, L. A. *et al.* An abundance of small exoplanets around stars with a wide range of metallicities. *Nature* **486**, 375–377 (2012).
13. Yi, S. *et al.* Toward better age estimates for stellar populations: the Y^2 isochrones for solar mixture. *Astrophys. J. Suppl. Ser.* **136**, 417–437 (2001).
14. Lovis, Ch. *et al.* The HARPS search for southern extra-solar planets. XXXI. Magnetic activity cycles in solar-type stars: statistics and impact on precise radial velocities. Preprint at <http://arxiv.org/abs/1107.5325> (2011).
15. Hatzes, A. P. *et al.* An investigation into the radial velocity variations of CoRoT-7. *Astron. Astrophys.* **520**, A93–A108 (2010).
16. Zechmeister, M. & Kuerster, M. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astron. Astrophys.* **496**, 577–584 (2009).
17. Howard, A. W. *et al.* A rocky composition for an Earth-sized exoplanet. *Nature* <http://dx.doi.org/10.1038/nature12767> (this issue).
18. Zeng, L. & Sasselov, D. A detailed model grid for solid planets from 0.1 through 100 Earth masses. *Publ. Astron. Soc. Pacif.* **125**, 227–239 (2013).
19. Rogers, L. A., Bodenheimer, P., Lissauer, J., Seager, S., The low density limit of the mass-radius relation for exo-Neptunes. *Bull. Am. Astron. Soc.* **43**, 402.04 (2011).
20. Léger, A., Rouan, D. & Schneider, J. Transiting exoplanets from the CoRoT space mission. VIII. CoRoT-7b: the first super-Earth with measured radius. *Astron. Astrophys.* **506**, 287–302 (2009).
21. Marcus, R. A., Sasselov, D., Hernquist, L. & Stewart, S. T. Minimum radii of super-Earths: constraints from giant impacts. *Astrophys. J.* **712**, L73–L76 (2010).
22. Lissauer, J. J. *et al.* All six planets known to orbit Kepler-11 have low densities. *Astrophys. J.* **770**, 131–145 (2013).
23. Charbonneau, D. *et al.* A super-Earth transiting a nearby low-mass star. *Nature* **462**, 891–894 (2009).
24. Smith, J. C. *et al.* Kepler presearch data conditioning II: a Bayesian approach to systematic error correction. *Publ. Astron. Soc. Pacif.* **124**, 1000–1014 (2012).

25. Stumpe, M. C. *et al.* Kepler presearch data conditioning I: architecture and algorithms for error correction in Kepler light curves. *Publ. Astron. Soc. Pacif.* **124**, 985–999 (2012).
26. Schwarz, G. E. Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).
27. Liddle, A. R. Information criteria for astrophysical model selection. *Mon. Not. R. Astron. Soc.* **377**, L74–L78 (2007).
28. Burnham, K. P. Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements This Letter was submitted simultaneously with the paper by Howard *et al.*¹⁷. Both papers are the result of a coordinated effort to carry out independent radial-velocity observations and studies of Kepler-78. Our team greatly appreciates the spirit of this collaboration, and we sincerely thank A. Howard and his team for the collegial work. We wish to thank the technical personnel of the Geneva Observatory, the Astronomical Technology Centre, the Smithsonian Astrophysical Observatory and the Telescopio Nazionale Galileo for their enthusiasm and competence, which made the HARPS-N project possible. The HARPS-N project was funded by the Prodex Program of the Swiss Space Office, the Harvard University Origins of Life Initiative, the Scottish Universities Physics Alliance, the University of Geneva, the Smithsonian Astrophysical Observatory, the Italian National Astrophysical Institute, the University of St Andrews, Queen's University Belfast and the University of Edinburgh. P.F. acknowledges support from the European Research Council/European Community through the European Union Seventh Framework Programme, Starting Grant agreement number 239953, and from the Fundação para a Ciência e a Tecnologia through grants PTDC/CTE-AST/098528/2008 and PTDC/CTE-AST/098604/2008. The research leading to these results received funding from the

European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement number 313014 (ETAEARTH).

Author Contributions The underlying observation programme was conceived and organized by F.P., A.C.C., D.W.L., C.L., D. Ségransan, S.U. and E.M. Observations with HARPS-N were carried out by A.C.C., A.S.B., D.C., R.C., C.D.D., X.D., P.F., A.F.M.F., S.G., A.H., R.D.H., M.L.-M., V.N., D. Pollacco, D.Q., K.R., A. Sozzetti, A. Szentgyorgyi and C.A.W. The data-reduction pipeline was adapted and updated by C.L., who also implemented the correction for charge-transfer-efficiency errors and the automatic computation of the activity indicator $\log(R'_{\text{HK}})$. M.L.-M. and S.G. independently computed the S-index and $\log(R'_{\text{HK}})$ values. A.C.C., D. Ségransan, A.S.B. and X.D. analysed the data using the offset-correction method. A.C.C. and D. Ségransan re-analysed the data for the determination of the stellar rotational period based on the Kepler light curve. P.F. investigated for possible correlations between the radial velocities and the line bisector. L.A.B. conducted the stellar parameter classification analysis for the re-determination of the stellar parameters based on HARPS-N spectra. An independent determination of the stellar parameters was conducted by L.M. by analysis of the cross-correlation function. D. Ségransan compared many different models to fit the observed data and selected the most appropriate by using the Bayesian information criterion. D. Ségransan performed a detailed MCMC analysis for the determination of the planetary parameters, with contributions also from A.S.B. and A. Sozzetti. F.P. was the primary author of the manuscript, with important contributions by D.C., D. Ségransan, D. Sasselov, C.A.W., K.R. and C.L. All authors are members of the HARPS-N Science Team and have contributed to the interpretation of the data and the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.P. (francesco.pepe@unige.ch).

METHODS

Photometric determination of stellar rotational period of Kepler-78. In ref. 8, the Kepler light curve of Kepler-78 was analysed and was de-trended using the PDC-MAP algorithm (Extended Data Fig. 2), which preserves stellar variability^{24,25}. The light curve displays clear rotational modulation with a peak-to-valley amplitude that varies between 0.5% and 1.5%, and a period of 12.6 ± 0.3 d. We confirmed the rotational period by computing the autocorrelation function of the PDC-MAP light curve: Using a fast Fourier transform we compute the power spectral density from which the autocorrelation function (ACF) is obtained using the inverse transform. We immediately derive a rotational period of 12.6 d (Extended Data Fig. 3a). The amplitudes of successive peaks decay on an e-folding timescale of about 50 d, which we attribute to the finite lifetimes of individual active regions. The power density distribution in Extended Data Fig. 3b finally shows a peak at the stellar rotational period as well as at its first and second harmonics. The main signal at period $P = 0.355$ d of Kepler-78b, as well as its harmonics, are easily identified at shorter periods.

HARPS-N observations and stellar parameters. To explore the feasibility of the programme, we performed five hours of continuous observations during a first test night in May 2013. This test night allowed us to determine the optimum strategy and to verify whether the measurement precision was consistent with expectations. Indeed, 12 exposures, each of 30 min, led to an observed dispersion of the order of 2.5 m s^{-1} , close to the expected photon noise. We therefore decided to dedicate six full HARPS-N nights to the observation of Kepler-78b in June 2013. Given the excellent stability of the instrument (typically less than 1 m s^{-1} during the night) and the faintness of the star, we observed without the simultaneous reference source^{10,11} that usually serves to track potential instrumental drifts. Instead, the second fibre of the spectrograph was placed on the sky to record possible background contamination during cloudy moonlit nights. Owing to excellent astrometric conditions, we gathered a total of 81 exposures, each of 30 min, free of moonlight contamination and with an average signal-to-noise ratio (SNR) of 45 per extracted pixel at a wavelength of $\lambda = 550 \text{ nm}$. An extracted pixel covers a wavelength bin of 0.000145 nm .

A first analysis of these observations confirmed the presence of the planetary signal. However, it also confirmed that, as suggested in ref. 8, the stellar variability induces radial-velocity variations much larger than the planetary signal, although on very different timescales. To consolidate our results and improve the precision of our planetary-mass measurement, we decided to perform additional observations during the months of July and August 2013. We preferred, however, to observe Kepler-78 only twice per night, around quadrature (at maximum and minimum expected radial velocity), to minimize observing time and to maximize the information on the amplitude. This strategy allowed us to determine the (low-frequency) stellar contribution as the sum of the two nightly measurements and the (high-frequency) planetary signal as the difference between them. We finally obtained a total of 109 high-quality observations over three months, with an average photon-noise-limited precision of 2.3 m s^{-1} .

The large number of high-SNR spectra gathered by HARPS-N allowed us to re-determine the stellar parameters using the stellar parameter classification (SPC) pipeline¹². Each high-resolution spectrum ($R = 115,000$) yields an average SNR per resolution element of 91 in the MgB region. The weighted average of the individual spectroscopic analyses resulted in final stellar parameters of $T_{\text{eff}} = 5058 \pm 50 \text{ K}$, $\log(g) = 4.55 \pm 0.1$, $[\text{M}/\text{H}] = -0.18 \pm 0.08$ and $v \sin(i) = 2 \pm 1 \text{ km s}^{-1}$, in agreement, within the uncertainties, with the discovery paper. The value for $v \sin(i)$ is, however, poorly determined by SPC. Therefore, we adopted an internal calibration based on the full-width at half-maximum of the cross-correlation functions to compute the projected rotational velocity, which yielded $v \sin(i) = 2.8 \pm 0.5 \text{ km s}^{-1}$. We note that, assuming spin-orbit alignment, the rotational velocity and our estimate of the stellar radius yield a rotational period of 13 d. This value is in agreement with the stellar rotational period determined from photometry.

The stellar parameters from SPC¹² have been input to the Yonsei-Yale stellar evolutionary models¹³ to estimate the mass and radius of the host star. We obtain $M_* = (0.758 \pm 0.046)M_{\odot}$ for the stellar mass and $R_* = 0.737^{+0.034}_{-0.042}R_{\odot}$ for the radius, in agreement, within the uncertainties, with the discovery paper. The Ca II HK activity indicator is computed by the online and automatic data-reduction pipeline, which gives an average value of $\log(R'_{\text{HK}}) = -4.52$ when using a colour index of $B-V = 0.91$ for Kepler-78. The stellar parameters are summarized in Extended Data Table 1.

Radial-velocity model selection. It is interesting to note that the signature of Kepler-78b can be retrieved despite the large stellar signals superimposed on the radial velocity induced by the planet. To demonstrate this, we adjusted the data with a simple model consisting of a cosine and the star's systemic velocity, while fixing the period and time of transit to the published values⁸. We compared the results of this simple model with a simple constant using the Bayesian information criterion^{26–28} (BIC). We derived the relative likelihood of the two models, also called the evidence ratio, to be $e^{-(1/2)\Delta\text{BIC}_i} = 4 \times 10^{-10}$. This first estimate tells us that our

cosine model is much superior to the simple constant. In other words, we can say that we have a clear detection of a signal of semi-amplitude $K_p = 1.88 \pm 0.47 \text{ m s}^{-1}$. Although certainly biased owing to the lack of stellar activity de-trending, the result provides a confirmation of the existence of Kepler-78b and a first estimation of its mass.

To model the stellar signature, we followed two different approaches. The first one consists of removing any stellar effect occurring on a timescale longer than 2 d by adjusting nightly offsets to the data. This method has the main advantage of not relying on any analytical model and it overcomes the difficulty of modelling non-stationary processes that often characterize stellar activity. The approach is also well suited to our problem because the period of the planet is very short. Its only drawback comes from the large number of additional parameters (21 offsets, one per night), which is a direct consequence of our observing strategy. The second approach consists of modelling the stellar activity as a set of sinusoids or Keplerian functions. This approach makes sense provided that spot groups and plagues are coherent on a timescale similar to the radial-velocity observation time span. For Kepler-78, the ACF of the light curve shows a 1/e de-correlation of ~ 50 d (Extended Data Fig. 3a), which compares well with the 97-d time span of the HARPS-N observations.

In total, we studied a series of more than 30 different models of different complexity. We have compared these models using the BIC²⁸ evidence ratio, ER, and the BIC weight, w_i , to find the best few models:

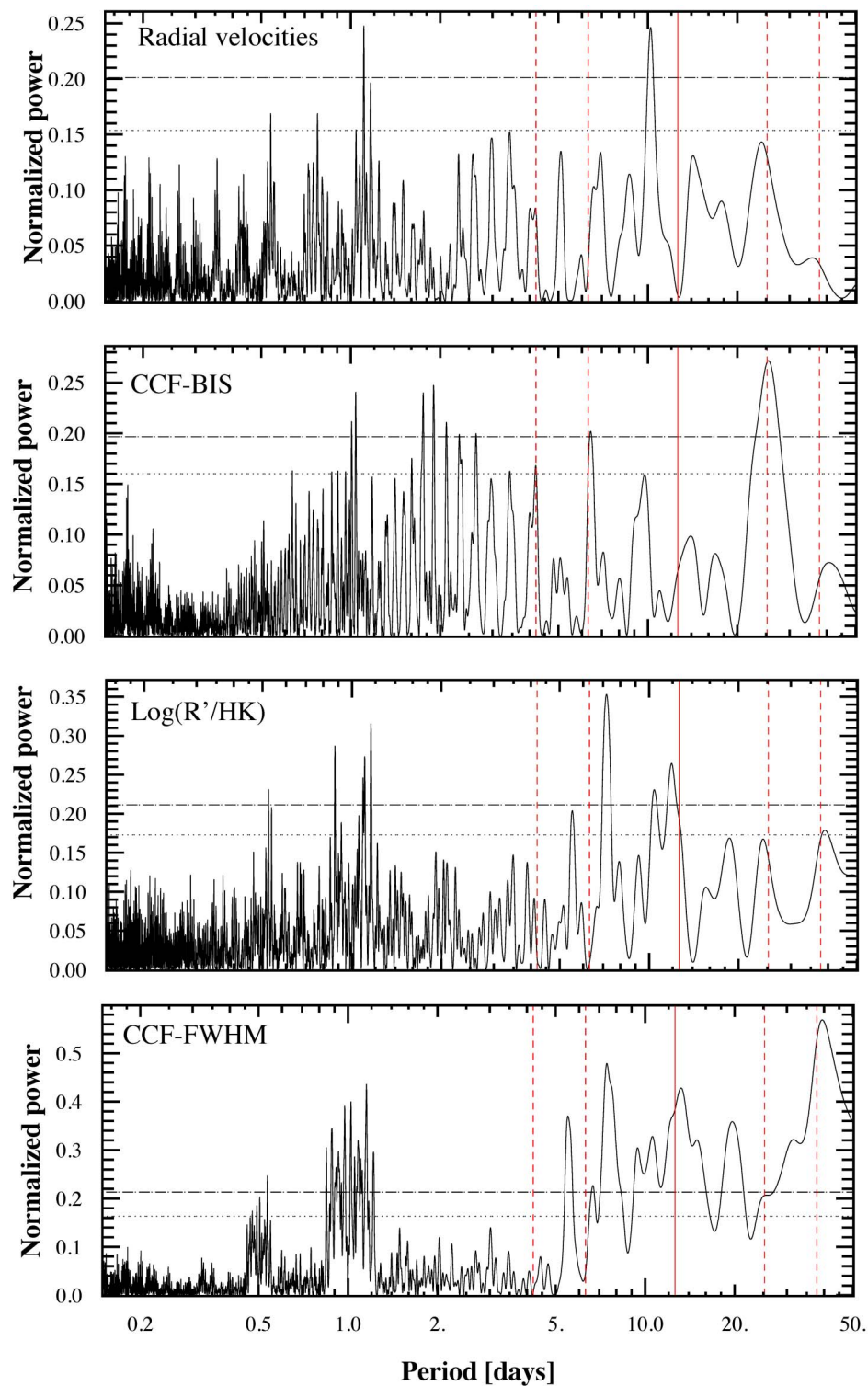
$$w_i = \frac{e^{-(1/2)\Delta\text{BIC}_i}}{\sum_j e^{-(1/2)\Delta\text{BIC}_j}}$$

Of all the models we considered, two are statistically much more significant. They consist of modelling the stellar activity as a sum of periodic signals. The best model, with a BIC weight of 0.71, predicts three Keplerians including the planetary signal ($P_1 = 0.355$ d, $P_2 = 4.2$ d and $P_3 = 10.04$ d; $e_1, e_2 = 0$). The second best contains four Keplerians ($P_1 = 0.355$ d, $P_2 = 4.2$ d, $P_3 = 6.5$ d and $P_4 = 23\text{--}58$ d; all eccentricities $e_i = 0$). De-trending of the stellar activity using nightly offsets shows much weaker evidence ratios. We therefore retained model 5 (Extended Data Table 2), which consists of one Keplerian describing the planet Kepler-78b and two additional 'signals', a sinusoid of period $P_2 = 4.2$ d and a slightly eccentric Keplerian with $P_3 = 10.04$ d. In Extended Data Table 3, we present the distribution of the parameters of the best model that results from our MCMC analysis (see below). Furthermore, Extended Data Fig. 4 shows the periodogram of the radial-velocity residuals after subtracting the stellar components. The planetary signal is now detected with a false-alarm probability significantly lower than 1%.

MCMC analysis. To retrieve the marginal distribution of the true mass of the planet and its density, we carry out an MCMC analysis based on the model selection process described in the previous section. We sample the posterior distributions using an MCMC with the Metropolis-Hastings algorithm. Because the model is very well constrained by the data, the MCMC starts from the solution corresponding to the maximum likelihood, and the MCMC parameter steps correspond to the standard deviation of the adjusted parameters. An acceptance rate of 25% is chosen. To obtain the best possible end result, we take as priors the transit parameters of Kepler-78b (ref. 8). Symmetric distributions are considered to be Gaussian, whereas asymmetric ones, such as that of the orbital inclination, are modelled by split-normal distributions using the published value of the mode of the distribution. We re-derive the radius of the planet using our improved stellar radius estimation and the planet-to-star radius ratio from the Kepler photometry⁸. All other parameters have uniform priors except for the period P , for which a modified Jeffrey's prior is preferred²⁹. We use $\sqrt{e} \cos(\omega)$ and $\sqrt{e} \sin(\omega)$ as free parameters, which translate into a uniform prior in eccentricity³⁰. The mean longitude, λ_0 , computed at the mean date of the observing campaign, is also preferred as a free parameter. It has the advantage of not being degenerate for low eccentricities, whereas our choice for the reference epoch, T_0 , reduces correlations between adjusted parameters. In this analysis, the MCMC has 2,000,000 iterations and converges after a few hundred iterations. The ACF of each parameter is computed to estimate the typical correlation length of our chains and to estimate a sampling interval to build the final statistical sample. All ACFs have a very short decay (1/e decay after 100 iterations and 1/100 decay after 300 iterations) and present no correlations on a larger iteration lag. We build our final sample using the 1/e-decay iteration lag, which is a good compromise between the size of the statistical sample and its de-correlation value. The final statistical samples consist of 20,000 elements, from which orbital elements and confidence intervals are derived. The resulting orbital elements for Kepler-78b are listed in Extended Data Table 3. The results for the mass, radius and density of the planet are given in Extended Data Table 4, and the distributions for the mass and the density are plotted in Extended Data Fig. 5. These distributions are smoothed for better rendering.

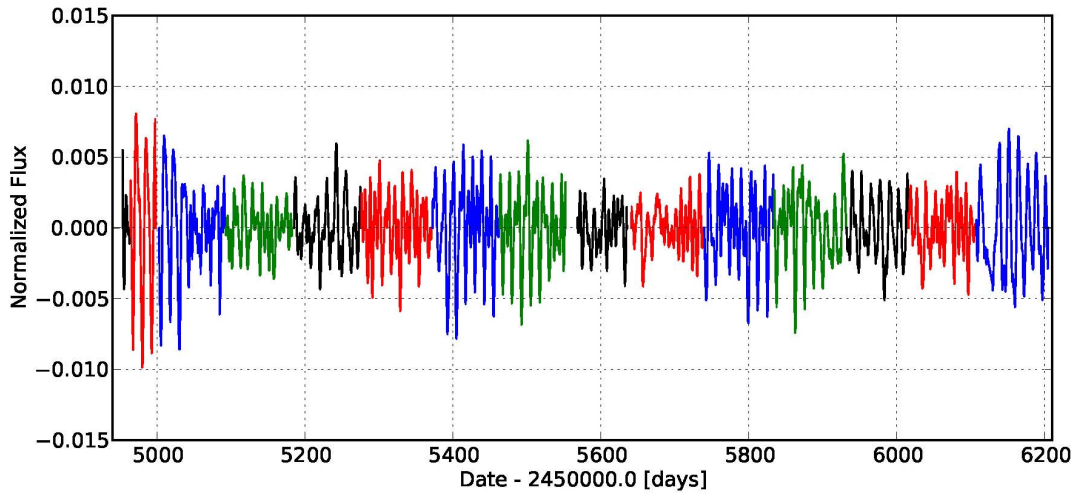
29. Gregory, P. C. *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge Univ. Press, 2005).

30. Anderson, D. R. et al. WASP-30b: a $61 M_{\text{Jup}}$ brown dwarf transiting a $V = 12$, F8 star. *Astrophys J.* **726**, L19–L23 (2011).

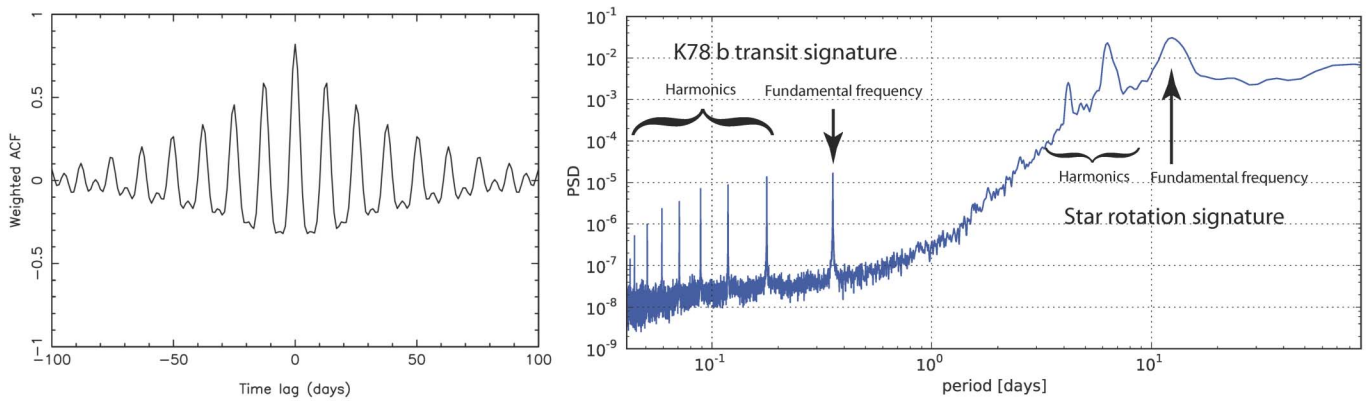


Extended Data Figure 1 | Generalized Lomb–Scargle periodogram of several parameters measured by HARPS-N. The panels show, from top to bottom, the periodogram of the radial velocities (RV) of Kepler-78, the line bisector (CCF-BIS), the activity indicator ($\log(R'_{\text{HK}})$) and the full-width at half maximum (CCF-FWHM) of Kepler-78. The dotted and dashed horizontal lines represent the 10% and 1% false-alarm probabilities, respectively. The

vertical lines show the stellar rotational period (solid) and its two first harmonics (dashed). All the indicators show excess energy at periods of around 6 d and above, indicating that the peak observed in the radial-velocity data at a period of about 10 d is most likely to have a stellar origin. The additional power in the line bisector periodogram at periods longer than 1 d is most probably induced by stellar spots.

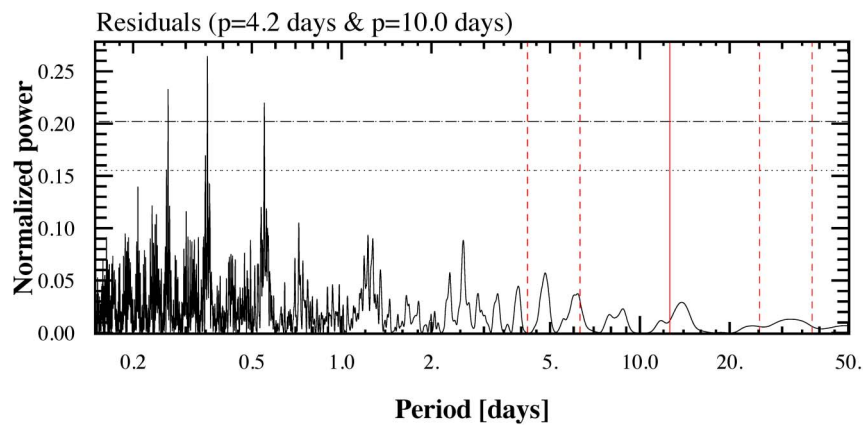


Extended Data Figure 2 | Kepler light curve of Kepler-78d. The data have been de-trended using the PDC-MAP algorithm. Different colours represent different quarters of observation.



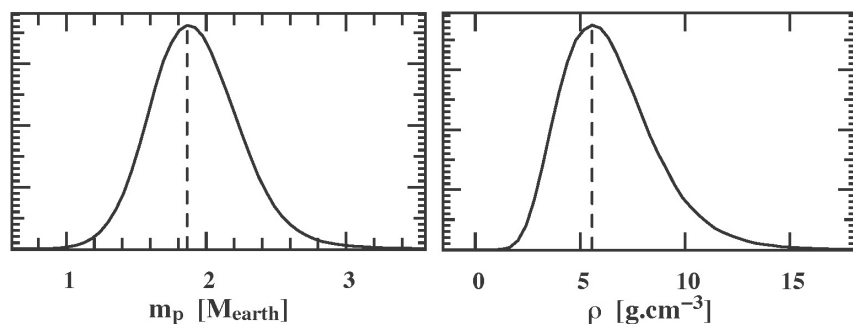
Extended Data Figure 3 | Spectral analysis of the Kepler light curve. Left panel, ACF of the Kepler light curve showing correlation peaks every 12.6 d and a decay on an e-folding timescale of ~ 50 d. Right panel, the power spectral distribution of the Kepler light curve. Peaks are well identified at the stellar

rotational period of 12.6 d and its two first harmonics. At shorter periods, the signal and several harmonics of the transiting planet Kepler-78b can be identified.



Extended Data Figure 4 | Periodogram of the radial-velocity residuals after subtraction of the 4.2-d and 10.0-d stellar components. The dotted and dashed horizontal lines represent the 10% and 1% false-alarm probabilities,

respectively. The signature of Kepler-78b (and its aliases) can now clearly be identified with a false-alarm probability significantly lower than 1%.



Extended Data Figure 5 | Probability density functions derived from the MCMC analysis. Probability density function of the planetary mass (left) and probability density function of the planetary density (right).

Extended Data Table 1 | Stellar parameters of Kepler-78 computed from HARPS-N spectra

Stellar parameter	Unit	Value	68.3% C.I.
*Effective temperature, T_{eff}	[K]	5058	± 50
*Surface gravity, $\log g$	[g in cm s ⁻²]	4.55	± 0.10
*Metallicity, [m/H]		-0.18	± 0.08
†Mass of the star, M_*	[M_{\odot}]	0.758	± 0.046
†Radius of the star, R_*	[R_{\odot}]	0.737	+0.034 -0.042
‡Ca II H&K activity indicator, $\log(R'_{\text{HK}})$		-4.52	$\pm 0.015^{\S}$
‡Projected rotational velocity, $v \sin i$	[km s ⁻¹]	2.8	± 0.5

* Obtained from an SPC analysis of the HARPS-N spectra. †Based on the Yonsei–Yale stellar evolutionary models¹³. ‡Output of the HARPS-N data-reduction pipeline using $B-V=0.91$. [§]The error indicates the standard deviation of the individual values.

Extended Data Table 2 | Comparison of the statistical ‘quality’ of all the considered models

Model	Description	Stellar Jitter	Planet	X^2	χ^2	N_{par}	ER	ΔER	w
1	D0	No	None	374.38	3.47	1	379.07	209.2	0.00
2	D0 + K1	No	K78 (fixed)	326.43	3.05	2	335.81	166.0	0.00
3	D0 + nightly offsets	21 offsets	K78 (fixed)	89.90	1.03	22	193.11	23.3	0.00
4	D0 + K2	2 Keplerians (P2=4.2; e2=0; P3=10.2; e3 adj)	None	166.03	1.69	11	217.63	47.8	0.00
5	D0 + K3	2 Keplerians (P2=4.2; e2=0; P3=10.2; e3 adj)	K78 (fixed)	122.94	1.24	10	169.85	0.00	0.710
6	D0 + K4	3 Keplerians (P2=4.2; e2=0; P3=6.5; e3 adj; P4=23.58; e4=0)	K78 (fixed)	120.40	1.23	11	172.00	2.2	0.242

Model 5 with the three Keplerians is clearly the one best representing the data, given the fact that its BIC evidence ratio, ER , is lowest and its BIC weight, w , highest. For comparison, the χ^2 and the reduced χ^2 of the residuals are also given. It is interesting to note that the most likely model does not necessarily have the lowest reduced χ^2 . N_{par} indicates the number of free parameters in the model. D0 represents a constant term, Kn is the number of Keplerians in the model. ei and Pi respectively indicate the eccentricity and the period of the i th planet.

Extended Data Table 3 | Orbital parameters (distributions) of the planet and parameters of the two additional Keplerians describing the star-induced signal as determined from the MCMC analysis

Parameters	Unit	Distribution mode	99% Confidence Interval
P_p	[days]	0.35500743	0.35500729 - 0.35500760
K_p	[m s ⁻¹]	1.986	1.243 - 2.679
e_p		0	fixed
λ_{0p}	[deg]	297.53	296.82 - 298.23
T_{0p}	[days]	1465.165144271	1465.165144235 - 1465.165144313
$m_p \sin(i)$	[M]	1.82	1.12 - 2.51
a_p	[10 ⁻³ AU]	8.95323	8.448 - 9.393
P_2	[days]	4.210	4.193 - 4.237
K_2	[m s ⁻¹]	3.98	2.66 - 5.29
e_2		0	fixed
λ_{02}	[deg]	146	118 - 166
P_3	[days]	10.037	9.977 - 10.272
K_3	[m s ⁻¹]	5.33	4.14 - 6.78
e_3		0.39	0.03 - 0.65
λ_{03}	[deg]	-320.0	-329 - -310
ω_3	[deg]	154	91 - 262

P_i represents the (orbital) period, K the semi-amplitude, e_i the eccentricity and λ_{0i} the mean longitude of the i th signal at the reference epoch, T_0 . The index i designates the planet (p) and the two additional (stellar) signals (2 and 3). $m_p \sin(i)$ is the minimum mass of the planet and a_p is the semi-major axis of the orbit.

Extended Data Table 4 | Planetary parameters derived from the MCMC analysis

Statistical parameter	Mass [M_\oplus]	Radius [R_\oplus]	Density [g cm ⁻³]
Mode	1.86	1.173	5.57
Median	1.91	1.194	6.13
68.3% confidence interval	1.61 – 2.24	1.084 – 1.332	4.26 – 8.59
99% confidence interval	1.17 – 3.00	0.942 – 1.584	2.34 – 14.29

The table gives the distributions of the mass, radius and density of Kepler-78b.

A rocky composition for an Earth-sized exoplanet

Andrew W. Howard¹, Roberto Sanchis-Ojeda², Geoffrey W. Marcy³, John Asher Johnson⁴, Joshua N. Winn², Howard Isaacson³, Debra A. Fischer⁵, Benjamin J. Fulton¹, Evan Sinukoff¹ & Jonathan J. Fortney⁶

Planets with sizes between that of Earth (with radius R_{\oplus}) and Neptune (about $4R_{\oplus}$) are now known to be common around Sun-like stars^{1–3}. Most such planets have been discovered through the transit technique, by which the planet's size can be determined from the fraction of starlight blocked by the planet as it passes in front of its star. Measuring the planet's mass—and hence its density, which is a clue to its composition—is more difficult. Planets of size $2–4R_{\oplus}$ have proved to have a wide range of densities, implying a diversity of compositions^{4,5}, but these measurements did not extend to planets as small as Earth. Here we report Doppler spectroscopic measurements of the mass of the Earth-sized planet Kepler-78b, which orbits its host star every 8.5 hours (ref. 6). Given a radius of $1.20 \pm 0.09R_{\oplus}$ and a mass of $1.69 \pm 0.41M_{\oplus}$, the planet's mean density of $5.3 \pm 1.8 \text{ g cm}^{-3}$ is similar to Earth's, suggesting a composition of rock and iron.

Kepler-78 is one of approximately 150,000 stars whose brightness was precisely measured at 30-min intervals for four years by NASA's Kepler spacecraft⁷. This star is somewhat smaller, less massive and younger than the Sun (Table 1). Every 8.5 hours the star's brightness declines by 0.02% as the planet Kepler-78b transits (passes in front of) the stellar disk. The planet's radius was originally measured⁶ to be $1.16^{+0.19}_{-0.14}R_{\oplus}$. Its mass could not be measured, although masses exceeding

$8M_{\oplus}$ could be ruled out because the planet's gravity would have deformed the star and produced brightness variations that were not detected.

We measured the mass of Kepler-78b by tracking the line-of-sight component of the host star's motion (the radial velocity) that is due to the gravitational force of the planet. The radial-velocity analysis is challenging not only because the signal is expected to be small (about $1–3 \text{ m s}^{-1}$) but also because the apparent Doppler shifts due to rotating star spots are much larger (about 50 m s^{-1} peak-to-peak). Nevertheless the detection proved to be possible, thanks to the precisely known orbital period and phase of Kepler-78b that cleanly separated the timescale of spot variations ($P_{\text{rot}} \approx 12.5$ days) from the much shorter timescale of the planetary orbit ($P \approx 8.5$ hours). We adopted a strategy of intensive Doppler measurements spanning 6–8 hours per night, long enough to cover nearly the entire orbit and short enough for the spot variations to be nearly frozen out.

We measured radial velocities using optical spectra of Kepler-78 that we obtained from the High Resolution Echelle Spectrometer (HIRES)⁸ on the 10-m Keck I Telescope. These Doppler shifts were computed relative to a template spectrum with a standard algorithm⁹ that uses a spectrum of molecular iodine superposed on the stellar spectrum as a reference for the wavelength scale and instrumental profile of HIRES

Table 1 | Kepler-78 system properties

Stellar properties	
Names	Kepler-78, KIC 8435766, Tycho 3147-188-1
Effective temperature, T_{eff}	$5,121 \pm 44 \text{ K}$
Logarithm of surface gravity, $\log[g (\text{cm s}^{-2})]$	4.61 ± 0.06
Iron abundance, $[\text{Fe}/\text{H}]$	$-0.08 \pm 0.04 \text{ dex}$
Projected rotational velocity, $V_{\text{sin i}}$	$2.6 \pm 0.5 \text{ km s}^{-1}$
Mass, M_{star}	$0.83 \pm 0.05 M_{\text{Sun}}$
Radius, R_{star}	$0.74 \pm 0.05 R_{\text{Sun}}$
Density, ρ_{star}	$2.8^{+0.7}_{-0.6} \text{ g cm}^{-3}$
Age	625 ± 150 million years
Planetary properties	
Name	Kepler-78b
Mass, M_{pl}	$1.69 \pm 0.41 M_{\oplus}$
Radius, R_{pl}	$1.20 \pm 0.09 R_{\oplus}$
Density, ρ_{pl}	$5.3^{+2.0}_{-1.6} \text{ g cm}^{-3}$
Surface gravity, g_{pl}	$11.4^{+3.5}_{-3.1} \text{ m s}^{-2}$
Iron fraction	0.20 ± 0.33 (two-component rock/iron model)
Orbital period, P_{orb} (from ref. 6)	$0.35500744 \pm 0.00000006$ days
Transit epoch, t_c (from ref. 6)	$2454953.95995 \pm 0.00015$ (BJD _{TDB})
Additional parameters	
$(R_{\text{pl}}/R_{\text{star}})^2$	217 ± 9 parts per million
Scaled semi-major axis, a/R_{star}	2.7 ± 0.2
Doppler amplitude, K	$1.66 \pm 0.40 \text{ m s}^{-1}$
Systemic radial velocity	$-3.59 \pm 0.10 \text{ km s}^{-1}$
Radial-velocity jitter, σ_{jitter}	$2.1 \pm 0.3 \text{ m s}^{-1}$
Radial-velocity dispersion	2.6 m s^{-1} (s.d. of residuals to best-fit model)

The stellar effective temperature and iron abundance were obtained by fitting stellar atmosphere models²¹ to iodine-free HIRES spectra, subject to a constraint on the surface gravity based on stellar evolution models²². We estimated the stellar mass and radius from empirically calibrated relationships between those spectroscopic parameters²³. The refined stellar radius led to a refined planet radius. Planet mass and density were measured from the Doppler analysis. The stellar age is estimated from non-detection of lithium in the stellar atmosphere (Extended Data Fig. 1), the stellar rotation period, and magnetic activity. See Methods for details. Parameter distributions are represented by median values and 68.3% confidence intervals. Correlations between transit parameters are shown in Extended Data Fig. 2. Barycentric Julian dates in barycentric dynamical time, BJD_{TDB}.

¹Institute for Astronomy, University of Hawaii at Manoa, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ²Department of Physics, and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Astronomy Department, University of California, Berkeley, California 94720, USA. ⁴Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ⁵Department of Astronomy, Yale University, New Haven, Connecticut 06510, USA. ⁶Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA.

(Supplementary Table 1). Exposures lasted 15–30 min depending on conditions and produced radial velocities with $1.5\text{--}2.0\text{ m s}^{-1}$ uncertainties. The time series of radial velocities spans 45 days, with large velocity offsets between nights due to star spots (Fig. 1). Within each night the radial velocities vary by typically $2\text{--}4\text{ m s}^{-1}$ and show coherence on shorter timescales.

We modelled the radial-velocity time series as the sum of two components. One component was a sinusoidal function representing orbital motion (assumed to be circular). The orbital period and phase were held fixed at the photometrically determined values; the only free parameters were the Doppler amplitude K , an arbitrary radial-velocity zero point, and a velocity ‘jitter’ term σ_{jitter} to account for additional radial-velocity noise. The second component of the model, representing the spot variations, was the sum of three sinusoidal functions with periods P_{rot} , $P_{\text{rot}}/2$, and $P_{\text{rot}}/3$. The amplitudes and phases of the sinusoids and P_{rot} were free parameters. All together there were ten parameters and 77 data points. Using a Markov Chain Monte Carlo (MCMC) method to sample the allowed combinations of the model parameters, we found $K = 1.66 \pm 0.40\text{ m s}^{-1}$, corresponding to $M_{\text{pl}} = 1.69 \pm 0.41 M_{\oplus}$ (Fig. 1). This planet mass is consistent with an independent measurement using the HARPS-N spectrometer¹⁰.

Several tests were performed to gauge the robustness of the spot model. First, we varied the number of harmonics, checking at each stage whether any improvement in the fit was statistically significant. The three-term model was found to provide noticeable improvement over one-term and two-term models, but additional harmonics beyond $P_{\text{rot}}/3$ did not provide significant improvement. Second, we used a different spot model in which the spot-induced variation was taken to be a linear function of time specific to each night. The constant and slope of each nightly function were free parameters. With this model we found $M_{\text{pl}} = 1.50 \pm 0.44 M_{\oplus}$, consistent with the preceding results (see Methods and Extended Data Fig. 3). The larger uncertainty can be attributed to the greater flexibility of this piecewise-linear spot model, which permits discontinuous and probably unphysical variations between consecutive nights.

Kepler-78b is now the smallest exoplanet for which both the mass and radius are known accurately (Fig. 2), extending the domain of such measurements into the neighbourhood of Earth and Venus. Kepler-78b is 20% larger than Earth and is 69% more massive, suggesting commonality with the other low-mass planets ($4\text{--}8 M_{\oplus}$) below the rock composition contour in Fig. 2b. They are all consistent with rock/iron compositions and negligible atmospheres.

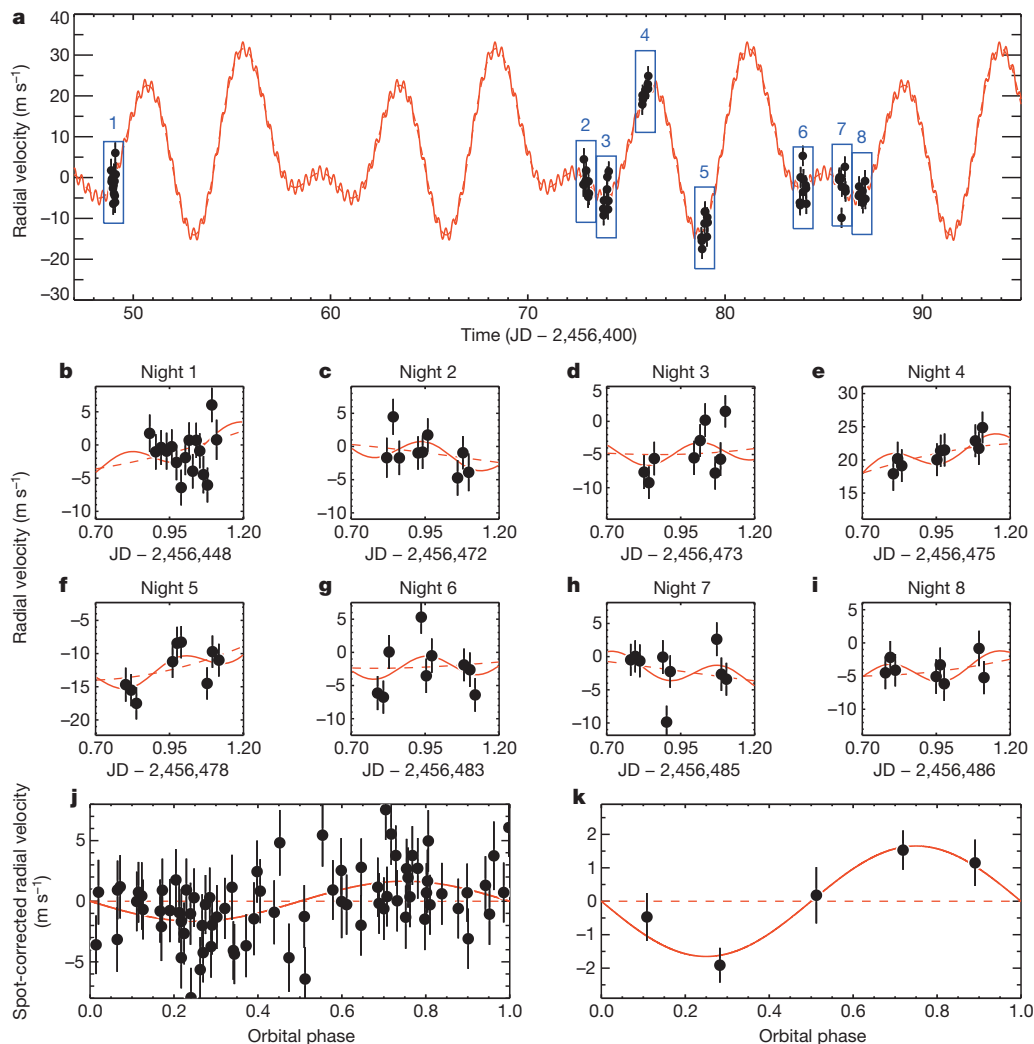


Figure 1 | Apparent radial-velocity variations of Kepler-78. **a**, The 38-day time series of relative radial velocities (black filled circles) from Keck-HIRES along with the best-fitting model (red line), with short-term variations due to orbital motion and long-term variations due to rotating star spots. Blue boxes identify the eight nights when high-cadence measurements were undertaken. JD, Julian day. **b–i**, For each individual night, the measured radial velocities (black filled circles), the spot + planet model (solid red lines) and spot model alone (dashed red lines) are shown. **j**, **k**, The phase-folded radial

velocities after subtracting the best-fitting spot model (**j**), and after binning in orbital phase and computing the mean radial velocities and s.e.m. for error bars (**k**). Planetary transits occur at zero orbital phase and solid red lines mark the phased planet model (in **j** and **k**). Each radial-velocity error bar in panels **a–j** represents the s.e.m. for the Doppler shifts of around 700 segments of a particular spectrum; it does not account for additional uncorrected radial-velocity ‘jitter’ from astrophysical and instrumental sources.

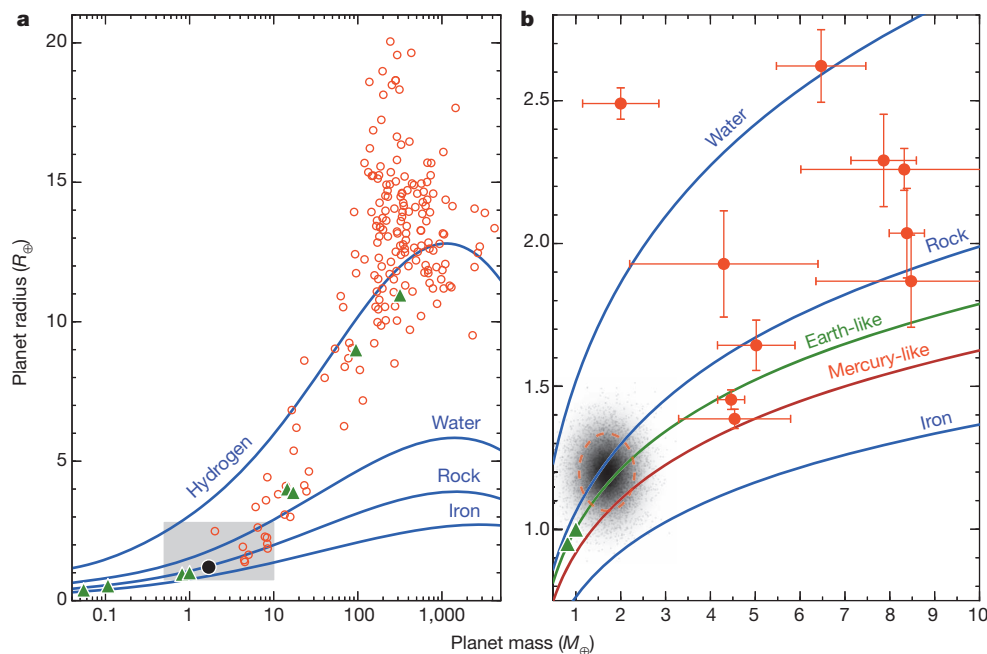


Figure 2 | Masses and radii of well-characterized planets. Extrasolar planets are denoted by red circles and Solar System planets are represented by green triangles. **a** spans the full range of sizes and masses on logarithmic axes. The shaded grey rectangle denotes the range of parameters shown in **b** on linear mass and radius axes. Kepler-78b is depicted as a black filled circle in **a** and as a distribution of allowed masses and radii with a dotted red ellipse marking the 68% confidence region in **b**. Model mass–radius relationships^{11,24} for

idealized planets consisting of pure hydrogen, water, rock (Mg_2SiO_4), and iron are shown as blue lines. Green and brown lines denote Earth-like composition (67% rock, 33% iron) and Mercury-like composition (40% rock, 60% iron). Exoplanet masses, radii and their associated errors are from the Exoplanet Orbit Database²⁵ (<http://exoplanets.org>; downloaded on 1 September 2013). Planets with fractional mass uncertainties of over 50% are not shown.

We explored some possibilities for the interior structure of Kepler-78b using a simplified two-component model¹¹ consisting of an iron core surrounded by a silicate mantle (Mg_2SiO_4). This model correctly reproduces the masses of Earth and Venus given their radii and assuming a composition of 67% silicate rock and 33% iron by mass. Applied to Kepler-78b, the model gives an iron fraction of $20\% \pm 33\%$, similar to that of Earth and Venus but smaller than that of Mercury (approximately 60%; ref. 12).

With a star–planet separation of 0.01 astronomical units (1 AU is the Earth-to-Sun distance), the dayside of Kepler-78b is heated to a temperature of 2,300–3,100 K. Any gaseous atmosphere around Kepler-78b would probably have been lost long ago to photoevaporation by the intense starlight¹³. However, based on the measured surface gravity of 11 m s^{-2} , the liquid and solid portions of the planet should be stable against mass loss of the sort¹⁴ that is apparently destroying the smaller planet KIC 12557548b (ref. 15).

Kepler-78b is a member of an emerging class of planets with orbital periods of less than half a day^{6,16,17}. Another member is KOI 1843.03 (refs 18 and 19), which has been shown to have a high density (more than about 7 g cm^{-3}), although the deduction in that case was based on a theoretical requirement to avoid tidal destruction rather than direct measurement. That planet's minimum density is similar to our estimated density for Kepler-78b (which is $5.3^{+2.0}_{-1.6} \text{ g cm}^{-3}$). These two planets provide a stark contrast to Kepler-11f, which has a similar mass to Kepler-78b, but a density that is ten times smaller²⁰.

With only a handful of low-mass planets with measured densities known (Fig. 2b), we see solid planets primarily in highly irradiated, close-in orbits and low-density planets swollen by thick atmospheres in somewhat cooler orbits. Measurements of additional planet masses and radii are needed to assess the significance of this pattern. Additional ultrashort-period planets with detectable Doppler amplitudes ($K \propto P^{-1/3}$) have been identified by the Kepler mission and are ready for mass measurements. With an ensemble of future measurements, the masses and radii of ultrashort-period planets may reveal a commonality or diversity of density and composition. This knowledge of hot solid planets may

be relevant to understanding the interiors of cooler extrasolar planets with atmospheres, establishing the range of core sizes in giant planet formation, and explaining Mercury's unusually high iron abundance.

METHODS SUMMARY

We fitted Keck-HIRES spectra of Kepler-78 with stellar atmosphere models using Spectroscopy Made Easy to measure the star's temperature, gravity and iron abundance. These spectroscopic parameters were used to estimate the host star's mass, radius and density—crucial parameters from which to determine the planet's mass, radius and density—from empirical relationships calibrated by precisely characterized binary star systems. Using this stellar density as a constraint, we reanalysed the Kepler photometry to refine the planet radius measurement. We observed Kepler-78 with HIRES using standard procedures including sky spectrum subtraction and wavelength calibration with a reference iodine spectrum. We measured high-precision relative radial velocities using a forward model where the de-convolved stellar spectrum is Doppler-shifted, multiplied by the normalized high-resolution iodine transmission spectrum, convolved with an instrumental profile, and matched to the observed spectra using a Levenberg–Marquardt algorithm that minimizes the χ^2 statistic. The time-series radial velocities on eight nights were analysed with several parametric models to account for the small-amplitude, periodic signal from the orbiting planet and the larger-amplitude, quasi-periodic apparent Doppler shifts that are due to rotating starspots. In our adopted harmonic spot model the star-spot signal was modelled as a sum of sine functions whose amplitudes and phases were free parameters. We sampled the multi-dimensional model parameter space with an MCMC algorithm to estimate parameter confidence intervals and to account for covariance between parameters. We found multiple families of models that described the data well and they gave consistent measures of the Doppler amplitude, which is proportional to the mass of Kepler-78b.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 September; accepted 11 October 2013.

Published online 30 October 2013.

- Howard, A. W. *et al.* Planet occurrence within 0.25 AU of solar-type stars from Kepler. *Astrophys. J.* **201** (Suppl.), 15 (2012).

2. Petigura, E., Marcy, G. W. & Howard, A. A plateau in the planet population below twice the size of Earth. *Astrophys. J.* **770**, 69 (2013).
3. Fressin, F. *et al.* The false positive rate of Kepler and the occurrence of planets. *Astrophys. J.* **766**, 81 (2013).
4. Valencia, D., Guillot, T., Parmentier, V. & Freedman, R. S. Bulk composition of GJ 1214b and other sub-Neptune exoplanets. *Astrophys. J.* **775**, 10 (2013).
5. Fortney, J. J. *et al.* A framework for characterizing the atmospheres of low-mass low-density transiting planets. *Astrophys. J.* **775**, 80 (2013).
6. Sanchis-Ojeda, R. *et al.* Transits and occultations of an Earth-sized planet in an 8.5-hour orbit. *Astrophys. J.* **774**, 54 (2013).
7. Borucki, W. J. *et al.* Kepler planet-detection mission: introduction and first results. *Science* **327**, 977–980 (2010).
8. Vogt, S. *et al.* HIRES: the high-resolution echelle spectrometer on the Keck 10-m telescope. *Proc. SPIE* **2198**, 362–375 (1994).
9. Butler, R. P. Attaining Doppler precision of 3 m s^{-1} . *Publ. Astron. Soc. Pacif.* **108**, 500–509 (1996).
10. Pepe, F. *et al.* An Earth-sized planet with an Earth-like density. *Nature* <http://dx.doi.org/10.1038/nature12768> (this issue).
11. Fortney, J. J. *et al.* Planetary radii across five orders of magnitude in mass and stellar insolation: application to transits. *Astrophys. J.* **659**, 1661–1672 (2007).
12. Hauck, S. A. *et al.* The curious case of Mercury's internal structure. *J. Geophys. Res.* **118**, 1204–1220 (2013).
13. Lopez, E. D., Fortney, J. J. & Miller, N. How thermal evolution and mass-loss sculpt populations of super-Earths and sub-Neptunes: application to the Kepler-11 system and beyond. *Astrophys. J.* **761**, 59 (2012).
14. Perez-Becker, D. & Chiang, E. Catastrophic evaporation of rocky planets. *Mon. Not. R. Astron. Soc.* **433**, 2294–2309 (2013).
15. Rappaport, S. *et al.* Possible disintegrating short-period super-Mercury orbiting KIC 12557548. *Astrophys. J.* **752**, 1 (2012).
16. Jackson, B., Stark, C. C., Adams, E. R., Chambers, J. & Deming, D. A survey for very short-period planets in the Kepler data. *Astrophys. J.* (submitted); preprint at <http://arxiv.org/abs/1308.1379> (2013).
17. Muirhead, P. S. *et al.* Characterizing the cool KOIs. III. KOI 961: a small star with large proper motion and three small planets. *Astrophys. J.* **747**, 144 (2012).
18. Ofir, A. & Dreizler, S. An independent planet search in the Kepler dataset. I. One hundred new candidates and revised Kepler objects of interest. *Astron. Astrophys.* **555**, A58 (2013).
19. Rappaport, S. *et al.* The Roche limit for close-orbiting planets: minimum density, composition, and application to the 4.2 hr planet KOI 1843.01. *Astrophys. J.* **773**, L15 (2013).
20. Lissauer, J. J. *et al.* All six planets known to orbit Kepler-11 have low densities. *Astrophys. J.* **770**, 131 (2013).
21. Valenti, J. A. & Fischer, D. A. Spectroscopic properties of cool stars (SPOCS). I. 1040 F, G, and K dwarfs from Keck, Lick, and AAT planet search programs. *Astrophys. J.* **159** (Suppl.), 141–166 (2005).
22. Dotter, A. *et al.* The Dartmouth stellar evolution database. *Astrophys. J.* **178** (Suppl.), 89–101 (2008).
23. Torres, G., Andersen, J. & Giménez, A. Accurate masses and radii of normal stars: modern results and applications. *Astron. Astrophys. Rev.* **18**, 67–126 (2010).
24. Seager, S. *et al.* Mass-radius relationships for solid exoplanets. *Astrophys. J.* **669**, 1279–1297 (2007).
25. Wright, J. T. *et al.* The exoplanet orbit database. *Publ. Astron. Soc. Pacif.* **123**, 412–422 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This Letter and another¹⁰ were submitted simultaneously and are the result of coordinated, independent radial-velocity observations and analyses of Kepler-78. We thank the HARPS-N team for their collegiality. We also thank E. Chiang, I. Crossfield, R. Kolbl, E. Petigura, and D. Huber for discussions, S. Howard for support, C. Dressing for a convenient packaging of stellar models, and A. Hatzes for a thorough review. This work was based on observations at the W. M. Keck Observatory granted by the University of Hawaii, the University of California, and the California Institute of Technology. We thank the observers who contributed to the measurements reported here and acknowledge the efforts of the Keck Observatory staff. We thank those of Hawaiian ancestry on whose sacred mountain of Mauna Kea we are guests. Kepler was competitively selected as the tenth Discovery mission with funding provided by NASA's Science Mission Directorate. J.N.W. and R.S.-O. acknowledge support from the Kepler Participating Scientist programme. A.W.H. acknowledges funding from NASA grant NNX12AJ23G.

Author Contributions This measurement was conceived and planned by A.W.H., G.W.M., J.A.J., J.N.W. and R.S.-O. HIRES observations were conducted by A.W.H., G.W.M., H.I., B.J.F. and E.S. The HIRES spectra were reduced and Doppler-analysed by A.W.H., G.W.M., H.I. and J.A.J. Data modelling was done primarily by A.W.H. and R.S.-O. A.W.H. was the primary author of the manuscript, with important contributions from J.N.W., R.S.-O. and J.J.F. Figures were generated by A.W.H., R.S.-O., B.J.F. and E.S. All authors discussed the results, commented on the manuscript and contributed to the interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.W.H. (howard@ifa.hawaii.edu) or R.S.-O. (rsanchis86@gmail.com).

METHODS

Stellar characterization. We fitted three Keck-HIRES spectra of Kepler-78 with stellar atmosphere models using Spectroscopy Made Easy (SME²⁶). The spectra have per-pixel signal-to-noise ratios of 220 at 550 nm. We used the standard wavelength intervals, line data and methodology²¹. Kepler-78 does not have a measured parallax with which to constrain luminosity and gravity. The initial analysis gave an effective temperature $T_{\text{eff}} = 5,119 \pm 44$ K, a gravity value of $\log[g \text{ (cm s}^{-2}\text{)}] = 4.751 \pm 0.060$, an iron abundance of $[\text{Fe}/\text{H}] = -0.054 \pm 0.040$ dex, and a projected rotational velocity of $V \sin i = 2.2 \pm 0.5$ km s⁻¹. These values are the mean of the SME results for the three spectra and the error bars are limited by systematics²¹. Because this combination of T_{eff} and $\log(g)$ is inconsistent with the Dartmouth stellar evolutionary model²², we recomputed stellar parameters with $\log(g)$ fixed at the value that is predicted by a stellar model at the value of T_{eff} from SME, resulting in the stellar parameters in Table 1. We note that the adopted $V \sin i = 2.6 \pm 0.5$ km s⁻¹ is consistent with an expectation based on a stellar rotation, size and an equatorial viewing geometry: $V \sin i \approx V_{\text{rot}} \approx 2\pi R_{\text{star}}/P_{\text{rot}} = 3.0$ km s⁻¹.

We estimated the stellar mass and radius using empirical relationships²³ based on non-interacting binary systems that parameterize R_{star} and M_{star} as functions of $\log(g)$, T_{eff} and $[\text{Fe}/\text{H}]$. We propagated the errors on the three SME-derived inputs to obtain $R_{\text{star}} = 0.74 \pm 0.05 R_{\text{Sun}}$ and $M_{\text{star}} = 0.83 \pm 0.05 M_{\text{Sun}}$. The M_{star} uncertainty comes from the 6% fractional scatter in the mass–radius relationship²³. We adopt these values when computing R_{pl} and M_{pl} . We checked for self-consistency of the empirical calibration by computing $\log(g)$ from the derived R_{star} and M_{star} , giving $\log[g \text{ (cm s}^{-2}\text{)}] = 4.62 \pm 0.06$.

As a consistency check, we explored two additional estimates of stellar parameters. First, the mass and radius from an evolutionary track in the Dartmouth model (one billion years, metal abundance $[\text{m}/\text{H}] = 0$) that match our adopted T_{eff} and $\log(g)$ values are $R_{\text{star}} = 0.77 \pm 0.04 R_{\text{Sun}}$ and $M_{\text{star}} = 0.85 \pm 0.05 M_{\text{Sun}}$. These values are consistent with our adopted results. Second, we used a recent study²⁷ of stellar angular diameters that parameterized R_{star} as a function of T_{eff} . This gives $R_{\text{star}} = 0.77 \pm 0.03 R_{\text{Sun}}$, where the uncertainty is the median absolute deviation on the calibration star radii.

We note that Kepler-78 has remarkably similar properties to the transiting planet host star HD 189733. These properties^{28,29} include $T_{\text{eff}} = 5,040$ K, $\log[g \text{ (cm s}^{-2}\text{)}] = 4.587$, $R_{\text{star}} = 0.76 R_{\text{Sun}}$ and $M_{\text{star}} = 0.81 M_{\text{Sun}}$, stellar activity index $\log(R'_{\text{HK}}) = -4.50$, and $P_{\text{rot}} = 11.9$ days. HD 189733 has spot-induced radial-velocity variations³⁰ of ~ 15 m s⁻¹ (rms).

The rotation period of Kepler-78 was previously measured to be 12.5 ± 1.0 days⁶. Using a relationship³¹ between age, mass and rotation period, we estimate its age to be 750 \pm 150 million years. The stellar age can also be estimated from the stellar magnetic activity measured by the S_{HK} index. We computed the spectral-type-independent activity index, $\log(R'_{\text{HK}})$, for all HIRES observations of this star and found a median value of -4.52 with a 1σ range of ± 0.03 . The computation made use of an estimated broadband photometric colour $B - V = 0.873$, converted²¹ from T_{eff} . This level of activity is consistent with the value for stars in the 625-million-year-old Hyades cluster³¹. We also constrained the age by searching for the age-sensitive Li I absorption line at 6,708 Å. Lithium is depleted relatively quickly in stars of this spectral type by convective mixing. Based on Li I measurements in three clusters with known ages³², our non-detection (Extended Data Fig. 1) suggests an age greater than around 500 million years. These three ages are self-consistent. We adopt an age of 625 million years with an approximate age uncertainty of 150 million years. We expect a star of this age and activity to have spots that cause radial-velocity variations at the >10 m s⁻¹ level.

Transit analysis. Transit parameters are crucial to estimating the planet radius, which in turn affects our ability to estimate the composition of the planet. These parameters were measured previously with the discovery of Kepler-78b (ref. 6). In that study the impact parameter b was nearly unconstrained because the 30-min time sampling of the Kepler long-cadence data cannot resolve the transit ingress time. This leads to an increased uncertainty on transit depth owing to the stellar limb-darkening profile. We constrained the transit parameters using the stellar density (ρ_{star}) obtained from the spectroscopic analysis. Assuming a circular orbit:

$$\rho_{\text{star}} = (3\pi/GP^2)(a/R_{\text{star}})^3$$

where a/R_{star} is the scaled semi-major axis, G is the gravitational constant, and P is the orbital period³³. This gives $a/R_{\text{star}} = 2.7 \pm 0.2$, a much tighter constraint than from the transit light curve alone ($a/R_{\text{star}} = 3.0^{+0.5}_{-1.0}$).

Aside from this additional constraint, our transit analysis is similar to the one in ref. 6. In brief, we analysed the Kepler long-cadence data from quarters 1–15 (a total of 3.7 years of nearly continuous observations) to construct a filtered, phase-folded light curve with a final cadence of 2 min. The light curve is modelled with a combination of a transit model³⁴, a model for the out-of-transit modulations and an occultation model. The most relevant transit parameters are the impact parameter,

the ratio of stellar radius to orbital distance, and the zero-limb-darkening transit depth. The model is calculated with a cadence of 15 s and averaged over the 30-min cadence of Kepler. In this new analysis, a/R_{star} is subjected to a Gaussian prior (2.7 ± 0.2), which leads to a well-measured impact parameter and a reduced uncertainty for the transit depth. We found the best-fit solution and explored parameter space using an MCMC algorithm. The final parameters are $(R_{\text{pl}}/R_{\text{star}})^2 = 217^{+9}_{-10}$ parts per million, impact parameter $b = 0.68^{+0.05}_{-0.02}$, orbital inclination $i = 75.2^{+2.6}_{-2.1}$ deg, a transit duration of 0.813 ± 0.014 hours, and $R_{\text{pl}} = 1.20 \pm 0.09 R_{\oplus}$. Error bars encompass 68.3% confidence intervals. Parameter correlations are plotted in Extended Data Fig. 2. These values are compatible with the previous estimate that was not constrained by ρ_{star} (ref. 6).

Radial-velocity measurements. We observed Kepler-78 with the HIRES echelle spectrometer⁸ on the 10-m Keck I telescope using standard procedures. Observations were made with the C2 decker (a rectangular opening in the HIRES entrance) (14×0.86 arcsec). This slit is long enough to simultaneously record spectra of Kepler-78 and the faint night sky. We subtracted the sky spectra from the spectra of Kepler-78 during the spectral reduction³⁵.

Light from the telescope passed through a glass cell of molecular iodine heated to 50 °C. The dense set of molecular absorption lines imprinted on the stellar spectra from 5,000 Å to 6,200 Å provide a robust wavelength scale against which Doppler shifts are measured, as well as strong constraints on the instrumental profile at the time of each observation^{36,37}. We also obtained three iodine-free ‘template’ spectra of Kepler-78 using the B3 decker (14×0.57 arcsec). These spectra were used to measure stellar parameters, as described above. One of them was de-convolved using the instrumental profile measured from spectra of rapidly rotating B stars observed immediately before and after. This de-convolved spectrum served as a ‘template’ for the Doppler analysis.

The HIRES observations span 45 days. On eight nights we observed Kepler-78 intensively, covering 6–8 hours per night. We also gathered a single spectrum on six additional nights to monitor the radial-velocity variations from spots. These once-per-night radial velocities were not used to determine the planetary mass and are shown in Extended Data Fig. 3 but not in Fig. 1.

We measured high-precision relative radial velocities using a forward model where the de-convolved stellar spectrum is Doppler-shifted, multiplied by the normalized high-resolution iodine transmission spectrum, convolved with an instrumental profile, and matched to the observed spectra using a Levenberg–Marquardt algorithm that minimizes the χ^2 statistic⁹. In this algorithm, the radial velocity is varied (along with nuisance parameters describing the wavelength scale and instrumental profile) until the χ^2 minimum is reached.

The times of observation (in heliocentric Julian days, HJD), radial velocities relative to an arbitrary zero point, and error estimates are listed in Supplementary Table 1 and plotted in Extended Data Fig. 3. Each radial-velocity error is the standard error on the mean radial velocity of about 700 spectral chunks (each spanning about 2 Å) that are separately Doppler-analysed. These error estimates do not account for systematic Doppler shifts from instrumental or stellar effects. We also measured the S_{HK} index for each HIRES spectrum. This index measures the strength of the inversion cores of the Ca II H and K absorption lines and correlates with stellar magnetic activity³⁸.

We measured the absolute radial velocity of Kepler-78 relative to the Solar System barycentre using telluric sky lines as a reference³⁹. The distribution of telluric radial velocities has a median value of -3.59 km s⁻¹ and a standard deviation of 0.10 km s⁻¹.

Harmonic radial-velocity spot model. Kepler-78 is a young active star, as demonstrated by the large stellar flux variations observed with Kepler. A previous study⁶ measured $P_{\text{rot}} = 12.5 \pm 1.0$ days using a Lomb–Scargle periodogram of the photometry. Inspection of the radial velocities measured over one month do indeed show some repeatability with a timescale of about 12–13 days, a sign that star spots are also inducing a large radial-velocity signal (see Extended Data Fig. 3). Using previous work⁴⁰, we modelled the radial-velocity signal induced by spots with a primary sine function at the rotation period of the star, followed by a series of sine functions representing the harmonics of the stellar rotation frequency. The planet-induced radial-velocity signal is modelled with a sinusoid, assuming zero eccentricity and using a linear ephemeris fixed to the best-fit orbital period and phase⁶. The final model for the radial velocity at time t is:

$$RV(t) = -K \sin[2\pi(t - t_c)/P] + \gamma + \sum_i a_i \sin[\varphi_i + i2\pi t/P_{\text{rot}}]$$

where K is the semi-amplitude of the planet-induced radial-velocity signal, t_c is a time of transit, P is the orbital period, γ is an arbitrary radial velocity offset, i runs from 1 to N , where N is the number of harmonics used, P_{rot} is the rotation period, and finally a_i and φ_i are the two parameters added for each of the N harmonics. The amplitude a_i is always chosen to be positive, and φ_i is constrained to be positive and smaller than 2π . The time t was set to zero at 2,456,446 in HJD format. P and t_c were

held fixed to their photometrically determined values (Table 1). K was free to take on positive and negative values to prevent a bias towards larger planet mass.

We used the Bayesian Information Criterion to choose the appropriate number of harmonics. This criterion states that for each additional model parameter the standard χ^2 function should decrease by at least $\ln(N_{\text{meas}})$ to be deemed statistically significant. In our case $N_{\text{meas}} = 77$ (the number of radial velocities on the eight nights with intensive observations). For each harmonic added, the best-fit χ^2 should decrease by at least 8.7 for the more complex model to be justified. The best-fit χ^2 values for $N = 1, 2, 3, 4$ and 5 were 1,822, 262, 163, 161 and 158, respectively. We used a Gaussian distribution as a prior on the rotation period in this analysis to select the number of harmonics. We adopted the $N = 3$ model because adding additional harmonics is not statistically justified.

This analysis does not rule out the possibility that non-consecutive harmonics provide a better fit to the data. We checked that for a model with three harmonics chosen from the first four, the first three harmonics is the best combination. We also checked that choosing only two harmonics out of the first four was never a better option than using the first three.

Using the spot model with the first three harmonics, we used an MCMC algorithm to explore the model parameter space. We added a radial-velocity ‘jitter’ term σ_{jitter} to account for the high value of the reduced χ^2 in the best-fit model without jitter (a value of 2.1), following a standard procedure⁴¹ to leave it as a free parameter. We maximized the logarithm of the likelihood function instead of minimizing the χ^2 function. We estimate the parameters describing the planet to be $K = 1.66 \pm 0.40 \text{ m s}^{-1}$, $\gamma = 4.4^{+2.7}_{-2.0} \text{ m s}^{-1}$, and $\sigma_{\text{jitter}} = 2.08^{+0.32}_{-0.29} \text{ m s}^{-1}$. The parameters describing the starspots are $P_{\text{rot}} = 12.78 \pm 0.04$ days, $a_1 = 3.6^{+3.6}_{-1.4} \text{ m s}^{-1}$, $a_2 = 10.5^{+3.3}_{-2.4} \text{ m s}^{-1}$, $a_3 = 10.2^{+1.5}_{-1.3} \text{ m s}^{-1}$, $\phi_1 = 4.4^{+1.2}_{-1.0}$, $\phi_2 = 3.9^{+0.2}_{-0.3}$ and $\phi_3 = 0.48 \pm 0.21$. These values are the median and 68.3% confidence regions of marginalized posterior distributions from the MCMC analysis. In this final run, P_{rot} was not subject to a prior and yet the value is compatible with the photometric estimate. Our estimate of K is inconsistent with zero at the 4σ level. This 4σ detection of Kepler-78b that is consistent with the orbital period and phase from Kepler gives us high confidence that we have detected the planet. The planet mass listed in Table 1 was computed from the values for K , P , i and M_{star} with the assumption of a circular orbit.

We searched for additional χ^2 minima to assess the sensitivity of our mass measurement to the spot model. We found a second family of solutions with $K = 1.85 \pm 0.43 \text{ m s}^{-1}$, $\sigma_{\text{jitter}} = 2.3 \pm 0.3 \text{ m s}^{-1}$, and $P_{\text{rot}} = 12.3$ days. Although our adopted solution with $P_{\text{rot}} = 12.8$ days is clearly preferred by the χ^2 criterion ($\chi^2 = 163$ versus 180), the broader model search demonstrates that our mass determination is relatively insensitive to details of the spot model.

We also calculated the K values with different combinations of two and three harmonics. For example, a model with the first, second and fourth harmonic gives $K = 1.78 \pm 0.45 \text{ m s}^{-1}$, with a slightly larger $\sigma_{\text{jitter}} = 2.3 \pm 0.3 \text{ m s}^{-1}$. A model with the first four harmonics gives $K = 1.77 \pm 0.41 \text{ m s}^{-1}$, with $\sigma_{\text{jitter}} = 2.2 \pm 0.3 \text{ m s}^{-1}$. Second, we included all of the radial velocities (including the six radial velocities measured on nights without intensive observations) and fitted the complete data set with three and four consecutive harmonics, giving $K = 1.80 \pm 0.43 \text{ m s}^{-1}$ and $\sigma_{\text{jitter}} = 2.4 \pm 0.3 \text{ m s}^{-1}$ for the three-harmonics model, and $K = 1.77 \pm 0.41 \text{ m s}^{-1}$ and $\sigma_{\text{jitter}} = 2.2 \pm 0.3 \text{ m s}^{-1}$ for the four-harmonics model. Although the coefficients and phases of the sine functions changed with each test, K remained compatible with the value from our adopted model. We also checked that K is not correlated with any other model parameters in the MCMC distribution.

We estimate the probability that radial-velocity noise fluctuations conspired to produce an apparently coherent signal with the precise period and phase of Kepler-78b to be approximately one in 16,000. This is the probability of a 4σ outlier for a normally distributed random variable. We adopt 4σ because the fractional error on K is approximately 4. Note that this is not the false alarm probability commonly computed for new Doppler detections of exoplanets. In those cases one must search over a wide range of orbital periods and phases to detect the planet, and also measure the planet’s mass. Here the existence of the planet was already well established⁶. Our task was to measure the planet’s mass given knowledge of its orbit. **Offset-slope radial-velocity spot model.** To gauge the sensitivity of our results to model assumptions, we considered a second radial-velocity model. Like the harmonic spot model, the offset-slope model consists of two components. The Doppler signal from the planet is a sinusoidal function of time with the period and phase held fixed at the values from the photometric analysis. The spot variations are approximated

as linear functions of time with slopes and offsets specific to each night, providing much greater model flexibility⁴². This model for the radial velocity at time t on night n is:

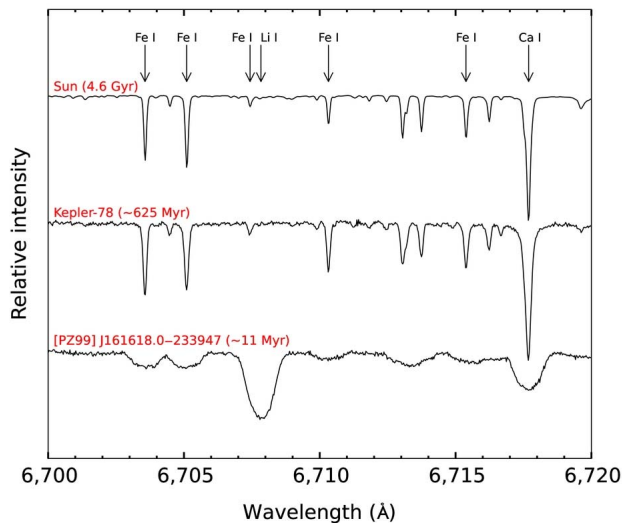
$$RV(t) = -K \sin[2\pi(t - t_c)/P] + \gamma_n + \dot{\gamma}_n(t - t_n)$$

where γ_n is a radial-velocity offset, $\dot{\gamma}_n$ is a radial-velocity slope (velocity per unit time), and t_n is the median time of observation specific to night n . The other symbols are as defined above. As with the previous model, a radial-velocity jitter term was added in quadrature to the errors and P and t_c were fixed (Table 1). All together, the offset-slope model contains 18 free parameters.

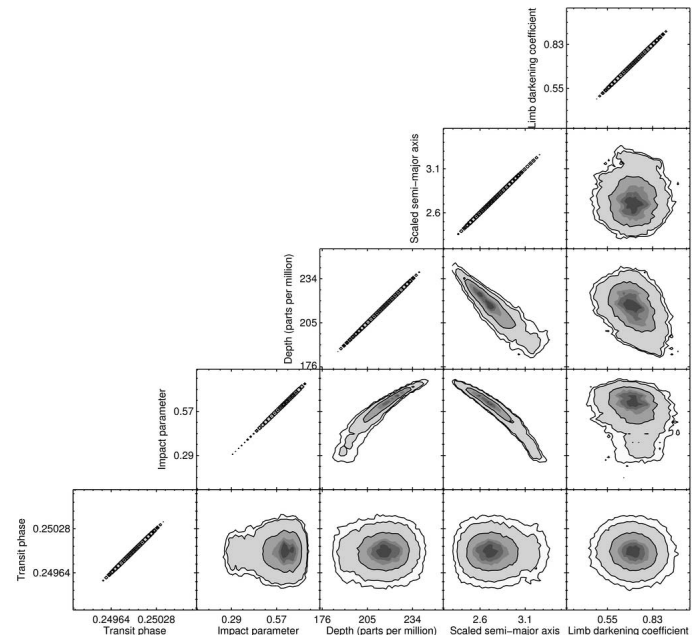
We used an MCMC algorithm to explore the model parameter space. The best-fit model and randomly selected models from the MCMC chain are shown in Extended Data Fig. 3. The key result is $K = 1.53 \pm 0.45 \text{ m s}^{-1}$, which is consistent with the value from the harmonic spot model. The lower precision of the offset-slope model (3.4σ versus 4.1σ significance) results from greater model flexibility. The slopes and offsets on nearby nights are not constrained to produce continuous spot variations as a function of time. For this reason we adopted the harmonic spot model.

As an additional test of the sensitivity to model details, we used the offset-slope framework to model a subset of the radial velocities. Within each night, we selected the median values from each group of three radial velocities ordered in time. This selection naturally rejects outlier radial velocities and matches the observing style on nights 2–8 when three groups of three measurements were made as close as possible to orbital quadratures (maximum or minimum radial velocity). (On night 8, the final group of radial velocities has only two measurements; we used the mean of those two radial velocities for this test.) Our MCMC analysis of these median radial velocities gave a similar result, $K = 1.26 \pm 0.38 \text{ m s}^{-1}$, that is consistent with the above results at about the 1σ level. We conclude that our detection of Kepler-78 is not strongly sensitive to spot model assumptions or to individual radial-velocity measurements.

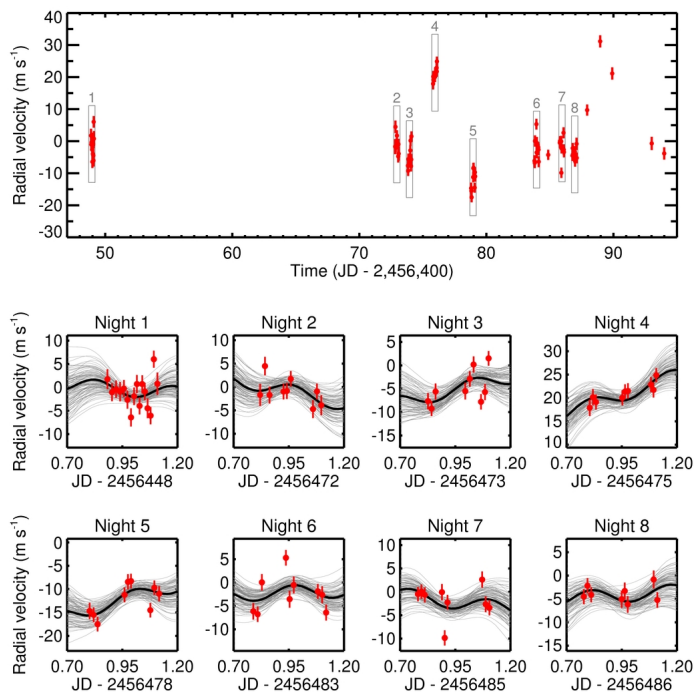
26. Valenti, J. & Piskunov, N. Spectroscopy made easy: a new tool for fitting observations with synthetic spectra. *Astron. Astrophys.* **118** (Suppl.), 595–603 (1996).
27. Boyajian, T. S. *et al.* Stellar diameters and temperatures. II. Main-sequence K- and M-stars. *Astrophys. J.* **757**, 112 (2012).
28. Torres, G., Winn, J. N. & Holman, M. J. Improved parameters for extrasolar transiting planets. *Astrophys. J.* **667**, 1324 (2012).
29. Henry, G. W. & Winn, J. N. The rotation period of the planet-hosting star HD 189733. *Astron. J.* **135**, 68–71 (2008).
30. Bouchy, F. *et al.* ELODIE metallicity-biased search for transiting hot Jupiters II. A very hot Jupiter transiting the bright K star HD 189733. *Astron. Astrophys.* **444**, L15–L19 (2005).
31. Mamajek, E. E. & Hillenbrand, L. A. Improved age estimation for solar-type dwarfs using activity-rotation diagnostics. *Astrophys. J.* **687**, 1264–1293 (2008).
32. Jones, B. F., Fischer, D. & Shetrone, M. The evolution of the lithium abundances of solar-type stars. VII. M34 (MGC 1039) and the role of rotation in lithium depletion. *Astron. J.* **114**, 352–362 (1997).
33. Seager, S. & Mallén-Ornelas, G. A unique solution of planet and star parameters from an extrasolar planet transit light curve. *Astrophys. J.* **585**, 1038–1055 (2003).
34. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
35. Batalha, N. *et al.* Kepler’s first rocky planet: Kepler-10b. *Astrophys. J.* **729**, 27 (2011).
36. Marcy, G. W. & Butler, R. P. Precision radial velocities with an iodine absorption cell. *Publ. Astron. Soc. Pacif.* **104**, 270–277 (1992).
37. Valenti, J. A., Butler, R. P. & Marcy, G. W. 1995, Determining spectrometer instrumental profiles using FTS reference spectra. *Publ. Astron. Soc. Pacif.* **107**, 966–976 (1995).
38. Isaacson, H. & Fischer, D. Chromospheric activity and jitter measurements for 2630 stars on the California planet search. *Astrophys. J.* **725**, 875–885 (2010).
39. Chubak, C. *et al.* Precise radial velocities of 2046 nearby FGKM stars and 131 standards. *Astrophys. J.* (submitted); available at <http://arxiv.org/abs/1207.6212> (2012).
40. Dumusque, X. *et al.* An Earth-mass planet orbiting α Centauri B. *Nature* **491**, 207–211 (2012).
41. Johnson, J. A. *et al.* Retired A stars and their companions. VII. 18 new jovian planets. *Astrophys. J.* **197** (Suppl.), 26 (2012).
42. Hatzes, A. P. *et al.* The Mass of CoRoT-7b. *Astrophys. J.* **743**, 75 (2011).
43. Peca, M. J., Mamajek, E. E. & Bubar, E. J. A revised age for upper Scorpius and the star formation history among the F-type members of the Scorpius-Centaurus OB association. *Astrophys. J.* **746**, 154 (2012).



Extended Data Figure 1 | Wavelength-calibrated spectra of three stars near the age-sensitive Li I line (6,708 Å). This line is not detected in the Kepler-78 spectrum, suggesting that lithium has been depleted, consistent with an age exceeding half a billion years for this K0 star. The lithium line is also not detected in the 4.6-billion-year-old Sun. (Gyr, billion years; Myr, million years.) It is clearly seen in the rotationally broadened spectrum of [PZ99] J161618.0 – 233947, a star whose spectral type (G8) is similar to that of Kepler-78, but that is much younger (about 11 million years)⁴⁴. Additional iron and calcium lines are labelled.



Extended Data Figure 2 | Correlations between model parameters in the transit analysis. Greyscale contours denote confidence levels, with thick black lines highlighting the 1 σ , 2 σ and 3 σ contour levels. The strongest correlations are between transit depth, scaled semi-major axis and impact parameter.



Extended Data Figure 3 | Apparent radial-velocity variations of Kepler-78 for the offset-slope model. The top panel shows the complete 45-day time series of relative radial velocities (red filled circles). Eight grey boxes highlight nights of intensive observations. The measurements from these nights are shown in the eight subpanels. In each subpanel, the radial velocities (red filled circles) and best-fit offset-slope model (solid black line) are shown. The radial-velocity curves for 100 randomly selected models from the MCMC chain are underplotted in grey, showing the range of variation within the model distribution.

Reducing the contact time of a bouncing drop

James C. Bird^{1*}, Rajeev Dhiman^{2*†}, Hyuk-Min Kwon^{2*} & Kripa K. Varanasi²

Surfaces designed so that drops do not adhere to them but instead bounce off have received substantial attention because of their ability to stay dry^{1–4}, self-clean^{5–7} and resist icing^{8–10}. A drop striking a non-wetting surface of this type will spread out to a maximum diameter^{11–14} and then recoil to such an extent that it completely rebounds and leaves the solid material^{15–18}. The amount of time that the drop is in contact with the solid—the ‘contact time’—depends on the inertia and capillarity of the drop¹, internal dissipation¹⁹ and surface–liquid interactions^{20–22}. And because contact time controls the extent to which mass, momentum and energy are exchanged between drop and surface²³, it is often advantageous to minimize it. The conventional approach has been to minimize surface–liquid interactions that can lead to contact line pinning^{20–22}; but even in the absence of any surface interactions, drop hydrodynamics imposes a minimum contact time that was conventionally assumed to be attained with axisymmetrically spreading and recoiling drops^{21,24}. Here we demonstrate that it is possible to reduce the contact time below this theoretical limit by using superhydrophobic surfaces with a morphology that redistributes the liquid mass and thereby alters the drop hydrodynamics. We show theoretically and experimentally that this approach allows us to reduce the overall contact time between a bouncing drop and a surface below what was previously thought possible.

Our experiments involve releasing a water drop (radius $R = 1.33$ mm, velocity $U = 1.2$ m s^{−1}) onto a superhydrophobic surface and filming the bounce dynamics with high-speed cameras (Fig. 1). The surface used

is a laser-ablated silicon wafer coated with fluorosilane, with chemical hydrophobicity and microscopic texture ensuring its superhydrophobic character (Fig. 1a inset). On this surface, the impacting drop viewed from the side (Fig. 1a) spreads to a nearly uniform film, retracts, and then lifts off within 12.4 ms. Simultaneously acquired top-view images show nearly axisymmetric dynamics throughout the process (Fig. 1b), consistent with past experiments^{15–18}. When the film is axisymmetric and uniformly thick, the edge retracts inward at a constant velocity and the centre remains stationary^{25,26} (Fig. 1c). This retraction velocity decreases with certain texture–liquid interactions (such as pinning), increasing the contact time^{20–22}. Theoretical models suggest that the shortest contact time is on a surface with the sparsest texture necessary to trap a thin layer of air^{21,22}. As this limit is approached, the drop dynamics become increasingly axisymmetric. Therefore, it is tacitly assumed that the minimum contact time should occur for a drop that recoils axisymmetrically with a centre that remains stationary until engulfed by the retracting rim.

We explored as an alternative non-axisymmetric recoil, or centre-assisted recoil. The basic idea is that if the hydrodynamics are altered such that the drop retracts with the liquid near the centre assisting with the recoil (Fig. 1d), contact time might be reduced further. To activate the drop centre, we propose adding designed macrotextures to the non-wetting surface to trigger a controlled asymmetry and non-uniform velocity field (Fig. 1d) in the retracting film. The combination of faster velocities in thinner film sections and smaller distances along certain directions should reduce contact time below the axisymmetric case.

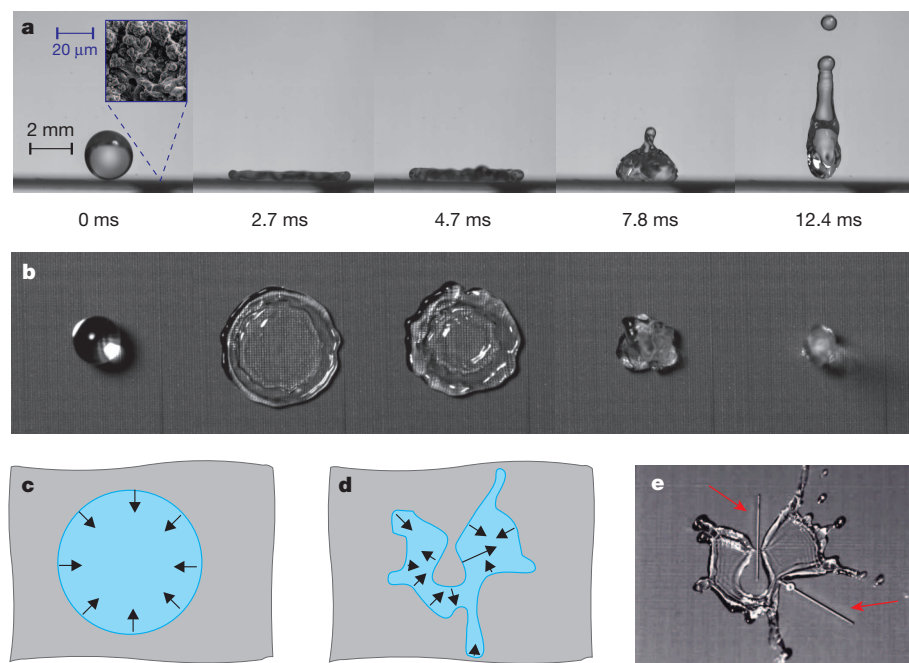


Figure 1 | A water drop bouncing on a superhydrophobic silicon surface. **a**, High-speed images of the bouncing show that the drop detaches from the surface after 12.4 ms (drop radius $R = 1.33$ mm; impact velocity $U = 1.2$ m s^{−1}). Inset, electron microscopy reveals the microscopic structure of the surface. **b**, Simultaneous top-view images demonstrate that the drop is nearly axisymmetric throughout the impact. **c**, The diagram portrays typical axisymmetric recoil with uniform retraction along the rim. **d**, A diagram portraying an arbitrary non-axisymmetric retraction in which the centre of the film assists in the recoil. **e**, Experimental evidence that such a recoil is possible when macrotexture (indicated by red arrows) is incorporated into the surface. For more details, see Supplementary Video 1.

¹Department of Mechanical Engineering, Boston University, Boston, Massachusetts 02155, USA. ²Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [†]Present address: 3M Purifications, Inc., 400 Research Parkway, Meriden, Connecticut 06033, USA.

*These authors contributed equally to this work.

The experimental demonstration of this concept uses an embossed macrotexture (indicated by red arrows in Fig. 1e) with an amplitude comparable to, but less than, the film thickness.

We promote non-axisymmetric centre-assisted recoil by spatially varying the film thickness, which can cause the retraction velocity to vary spatially. If the thickness h of the flattened drop were uniform, the rim should retract axisymmetrically with speed^{25,26} $V = \sqrt{2\gamma/\rho h}$ (Fig. 2a, left), where γ is the liquid–air surface tension and ρ is the liquid density. However, if the thickness were not uniform, the retraction velocity would be faster in the thinner regions with less mass to accelerate (Fig. 2a, right). As the faster retracting fronts move along the peak of the macrotexture (ridge), the centre opens, fragmenting the drop and decreasing the distance and time required for recoil.

We accordingly fabricated a superhydrophobic surface with two distinct length scales (Fig. 2b). The smaller length scale consists of hierarchical micrometre-scale and nanometre-scale features identical to those used in Fig 1a, b, imparting superhydrophobicity with minimal pinning.

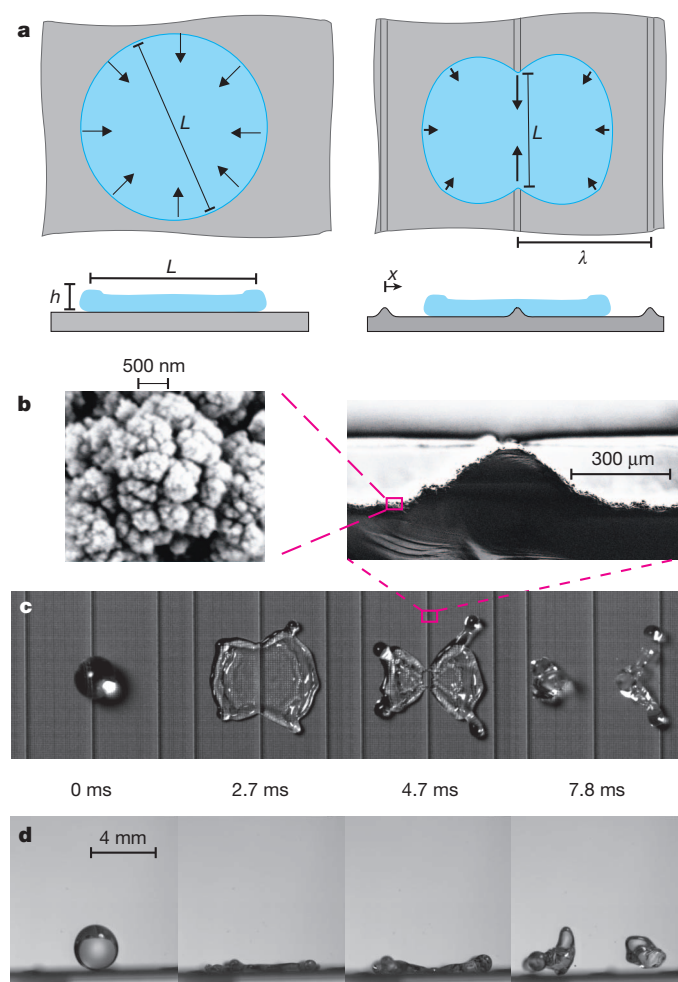


Figure 2 | Non-axisymmetric recoil can shorten contact time. **a**, The retraction speed of a film increases with decreasing thickness. Left and right, diagrams of a macroscopically textured surface and a macroscale textured surface, respectively; top and bottom, top-view and side-view diagrams illustrating how macroscale texturing can modify the thickness profile of the drop, leading to variations in recoil speed (indicated by the length of the arrows). **b**, As shown in these SEM images, we have fabricated a silicon surface with both submicrometre roughness and structure on a macroscopic ($\sim 100 \mu\text{m}$) scale by laser ablation. **c**, When a drop impacts the surface with the macroscopic structure, it moves rapidly along the ridge as it recoils. **d**, Simultaneous high-speed images captured from the side reveal that the overall contact time is reduced by 37% to 7.8 ms. For more details, see Supplementary Videos 2, 3.

The larger length scale consists of macroscopic features approaching the length scale of the film thickness h (Fig. 2b) for modifying the retraction hydrodynamics. The macrotexture height z varies as $z = a \sin^n(x/\lambda)$, where x is the horizontal distance, $a = 150 \mu\text{m}$, $n = 100$ and $\lambda = 4 \text{ mm}$ (see Methods).

Top-view images of a drop recoiling on the macrotexture show faster retraction along the ridge than in other directions (Fig. 2c). This variation in speed breaks the radial symmetry of the recoiling film, causing the liquid to move rapidly inward along the ridge such that more of the film participates in the recoil. Note that the drop is not split before impact, but divides during recoil as a result of the modified hydrodynamics. Synchronized side-view images of this drop (Fig. 2d) verify that the overall contact time is less than that on the same surface without the macrotextures (Fig. 1b).

Previous experiments indicate that the drop contact time t_c is independent of the dimensionless Weber number, We ($\equiv \rho U^2 R / \gamma$); instead, it scales with the inertial–capillary timescale^{1,21}, $\tau \equiv \sqrt{\rho R^3 / \gamma}$. To enable comparisons, we therefore report our contact times relative to τ . The minimum contact time for low-deformation impact ($We < 1$) can be approximated by the lowest-order oscillation period for a spherical drop²⁷, $t_c / \tau = \pi / \sqrt{2} \approx 2.2$. For large-deformation impact ($We > 1$), the contact time is similar even though the dynamics are distinctly different¹. Indeed, to the best of our knowledge, every past experiment documenting a drop bouncing on passive surface—including Leidenfrost drops²⁸—has reported a contact time greater than $t_c / \tau = 2.2$ (Extended Data Table 1), which translates to between 12 and 13 ms in our experiments.

A typical way to convey drop impact dynamics is to plot the radius of the wetted area as a function of time (Fig. 3a). Because of symmetry about the ridge, we find it most instructive to track the motion of the film perpendicular to the ridge, using the same axis when tracking the drop on the control surface. Inspection of the dynamics on the control surface (filled red squares in Fig. 3a) indicates that the drop first spreads to 2.5 times its initial radius and then recoils at a nearly constant rate, slowing down slightly when the flattened drop can no longer be approximated by a thin film ($r/R \approx 1$). At dimensionless time $t/\tau = 2.2$, the wetting radii in opposing directions contact, and the drop leaves the control surface.

The dynamics for the macrotextured surface are slightly more complex. The drop initially spreads over a time $T_s = 0.63$ and then begins to recoil (black filled circles in Fig. 3a). During the next time interval T_1 , the film recoils along the ridge faster than it recoils perpendicular to the ridge, splitting into two drop fragments (Fig. 2c). At this point, the outer rim of the initial drop continues to recoil inward while the newly formed inward rim recoils outward. This combined inward and outward recoil continues over the time interval T_2 . At dimensionless time $t/\tau = 1.3$, one of the fragments lifts off the surface and at $t/\tau = 1.4$, the remaining fragment lifts off. We denote the difference in contact time on the two surfaces as ΔT .

One might be tempted to rationalize this reduction, ΔT , by modifying the radius in the theoretical scaling to reduce the drop volume by half. However, this approach is not physically appropriate because the drop splits after it has spread out (Fig. 2c). Therefore, the film thickness depends on the initial radius, as opposed to the reduced radius (Supplementary Information, Extended Data Fig. 7). A better approach is to estimate ΔT using a hydrodynamic model that combines thin-film retraction, conservation of mass, and variations in film thickness due to the macrotexture. First, we note that the axisymmetric dimensionless retraction time on the control surface can be expressed as $T_r = T_1 + T_2 + \Delta T = r_{\text{max}} / V\tau$, where r_{max} is the maximum wetting radius and V is the average retraction velocity. Next, we approximate the ridge dewetting time as $T_1 \approx r_{\text{max}} / (V_p \tau)$, where V_p is the retraction velocity along the peak of the macrotexture. Finally, we estimate the interval over which the fragmented drops retract as $T_2 \approx (r_{\text{max}} - VT_1 \tau) / (2V\tau)$. Here we have assumed the velocity of the outward rim and the newly-formed inward rim to be equal to each other and to the velocity of the axisymmetric control film. This

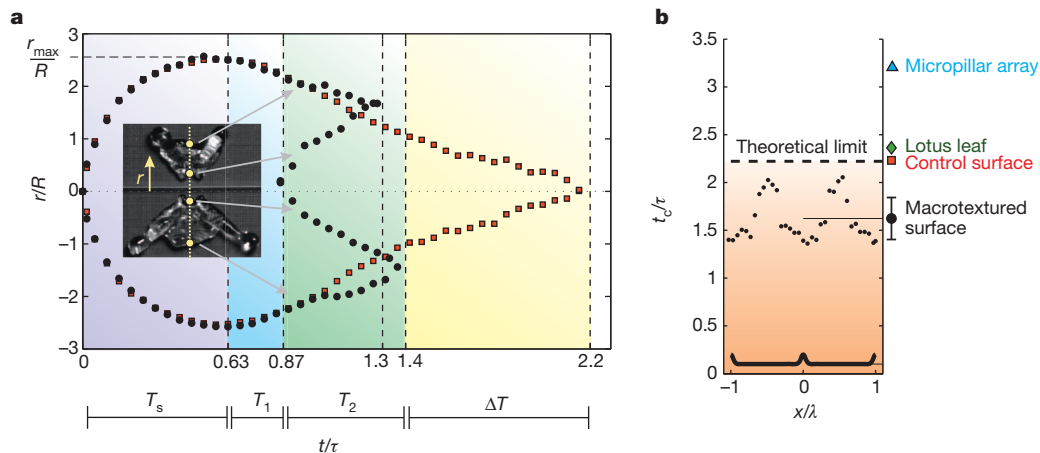


Figure 3 | The effect of macrotexture on drop impact dynamics and contact time. **a**, Plot of the contact line position (r ; see inset) of a water drop impacting the control surface in Fig. 1 (red squares) and the macrotextured surface in Fig. 2 (black circles). The shaded regions highlight the various timescales (T_s , T_1 , T_2 , ΔT) relevant to our model. See text for details. **b**, Unlike the solely micro-nanotextured surfaces, the contact time of a drop on the macrotextured surface (indicated by black dots) depends on where it lands along the periodic macrotexture (indicated by the thick line at the bottom). The average

contact time over the entire surface is shorter than that of non-macrotextured surfaces (error bars denote one standard deviation). The filled symbols depict the contact time on various superhydrophobic surfaces, including a micropillar array, a lotus leaf, and a control surface. ‘Theoretical limit’ refers to $t_c/\tau = 2.2$ as discussed in the text. The elapsed time t and contact time t_c are made non-dimensional by dividing by $\tau = \sqrt{\rho R^3/\gamma}$, the radius r is normalized by the drop radius R , and the distance along the surface x is normalized by the macrotexture wavelength λ .

approximation is reasonable, given our expectation of nearly uniform film thickness ($\sim h$), with the exception of the top of the ridge. Thus the thin-film retraction speed away from the ridge is approximately $V \approx \sqrt{2\gamma/(\rho h)}$, and the speed on the macrotexture peak is $V_p \approx \sqrt{(2\gamma)/[\rho(h-a)]}$, where a is the macrotexture amplitude. After noting that mass conservation requires $(4/3)\pi R^3 \rho \approx \pi r_{\max}^2 h \rho$, the previous expressions combine to reveal that $\Delta T \approx \frac{\sqrt{6}}{6} \left(1 - \sqrt{1 - \frac{a}{h}}\right)$. If there is no macrotexture ($a = 0$), then there is no contact time reduction ($\Delta T = 0$). If the macrotexture amplitude is equal to or greater than the film thickness ($a = h$), then the hydrodynamic model predicts a contact time reduction of $\Delta t_c \approx 0.4\tau$.

As Fig. 3 reveals, the model provides the correct order of magnitude, but underestimates the actual reduction by a factor of ~ 2 . This difference is due to assumptions that are visible in Fig. 3a. First, the retraction velocity is slower than predicted^{25,26} when the thin-film assumption breaks down. Second, the velocities of the inner and outer fronts are different, because the film thickness is not uniform. Last, the film away from the ridge spreads out further than the film on the ridge (Fig. 2c), resulting in an over-prediction of T_1 and under-prediction of ΔT . Nevertheless, the model elucidates the mechanism that reduces the overall contact time.

Careful inspection of Fig. 3a reveals that the two fragments leave the surface at slightly different times because the drop impacts the ridge slightly off-centre. At larger deviations from the ridge, this difference between the fragment lift-off times is more pronounced, increasing the overall contact time. The dimensionless contact times t_c/τ are reported for various landing locations along the periodic macrotexture x/λ (Fig. 3b). The contact time is shortest when the drop impacts directly on the ridge, increasing as the drop lands further away from the ridge, and then decreasing as the drop approaches the next ridge. By averaging over the wavelength, we find that the mean contact time over the entire surface is $t_c/\tau = 1.6$ with standard deviation $\sigma = 0.2$, a time significantly shorter than that on the control surface (Fig. 3b). For comparison, a drop under identical conditions contacted a lotus leaf for $t_c/\tau = 2.3$ and a micropillar array for $t_c/\tau = 3.2$ (Fig. 3b; Supplementary Video 3).

To confirm that the reduction in contact time is a result of the macrotexture geometry, and therefore a general phenomenon, we fabricated similar macrotextures in aluminium and copper by milling

ridges followed by microtexturing and coating with fluorosilane (Fig. 4a, b; Extended Data Figs 4, 5). The recoil dynamics are similar to those obtained on the macrotextured laser-ablated silicon surface (Fig. 2c; Fig. 4a, b).

We have searched to see if a similar system might exist in nature. We discovered that both the wings of the Morpho butterfly (*Morpho didius*) and the leaves of the nasturtium plant (*Tropaeolum majus* L.) have multiple superhydrophobic ridges, or veins, on a similar scale to our macrotextured surfaces. We find that centre-assisted recoil extends to

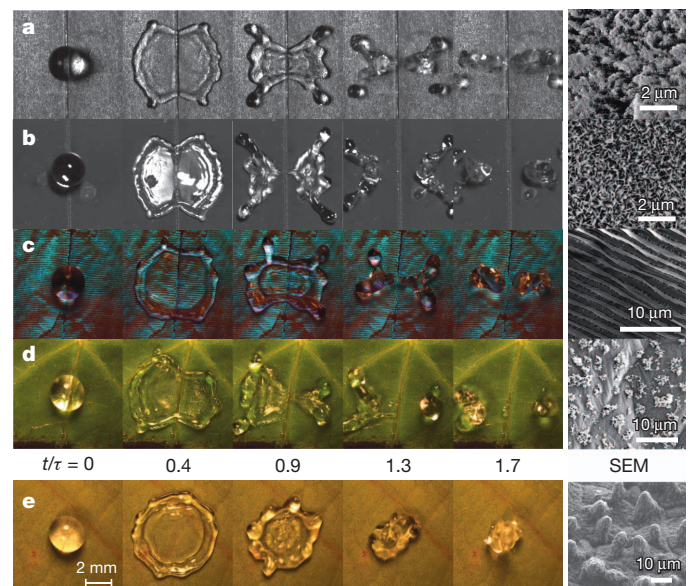


Figure 4 | Recoil dynamics generalize to a wide range of materials and microtextures. The figure shows top-view images of droplets impacting various surfaces; the SEM images of their respective microtextures are shown in the rightmost column. **a**, Anodized aluminium oxide with a milled macroscopic texture, pitted microtexture and a fluorinated coating. **b**, Etched copper oxide with a milled macroscopic texture, spiked microtexture and a fluorinated coating. **c**, A vein on the wing of a Morpho butterfly (*M. didius*). **d**, A vein on a nasturtium leaf (*T. majus* L.). **e**, For comparison, the same drop on a lotus leaf exhibits axisymmetric recoil. For all cases, $We = 30$. For more details, see Supplementary Video 4.

these surfaces as well (Fig. 4c, d; Extended Data Fig. 6), and that the overall drop contact time is significantly reduced from that of impact on macroscopically smooth surfaces, such as the lotus leaf, often considered the 'gold standard' of superhydrophobic surfaces (Fig. 4e). Further studies are needed to determine if there are any advantages for certain biological surfaces to contain these structures.

Although we have focused on a specific macrotexture that creates two distinct retraction velocities, centre-assisted recoil can occur for other macrotextures that modify the retraction hydrodynamics (such as Fig. 1e). These surfaces can be designed to reduce the drop contact time relative to other significant timescales, such as freezing. Indeed, molten tin drops impacting such surfaces are able to bounce off the surface before solidification (Supplementary Information, Extended Data Figs 1, 2, 3) and we expect that this approach could be extended to surfaces exposed to freezing rain to prevent icing. The new class of non-wetting surfaces that we present here could be useful for applications where staying dry under drop impingement is beneficial^{8–10,29,30}.

METHODS SUMMARY

The contact time of bouncing drops was obtained from the sequence of simultaneous top- and side-view images of drop impact captured by two high-speed cameras. The control and macrotextured ridge silicon surfaces were fabricated by ablating the surface using a Nd:YAG laser. The textured aluminium surface (Extended Data Fig. 4) was fabricated by milling ridges in aluminium and then performing a two-step anodization process consisting of polishing and etching. The textured copper oxide surface (Extended Data Fig. 5) was fabricated by milling ridges in copper and then treating with sodium hydroxide solution. All of the surfaces were coated with fluorosilane to render them superhydrophobic.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 June; accepted 26 September 2013.

- Richard, D., Clanet, C. & Quéré, D. Contact time of a bouncing drop. *Nature* **417**, 811 (2002).
- de Gennes, P. G., Brochard-Wyart, F. & Quéré, D. *Capillarity and Wetting Phenomena* (Springer, 2004).
- Gao, X. F. & Jiang, L. Water-repellent legs of water striders. *Nature* **432**, 36 (2004).
- Quéré, D. Non-sticking drops. *Rep. Prog. Phys.* **68**, 2495–2532 (2005).
- Blossey, R. Self-cleaning surfaces — virtual realities. *Nature Mater.* **2**, 301–306 (2003).
- Tuteja, A. *et al.* Designing superoleophobic surfaces. *Science* **318**, 1618–1622 (2007).
- Deng, X., Mammen, L., Butt, H. J. & Vollmer, D. Candle soot as a template for a transparent robust superamphiphobic coating. *Science* **335**, 67–70 (2012).
- Mishchenko, L. *et al.* Design of ice-free nanostructured surfaces based on repulsion of impacting water droplets. *ACS Nano* **4**, 7699–7707 (2010).
- Meuler, A. J., McKinley, G. H. & Cohen, R. E. Exploiting topographical texture to impart icephobicity. *ACS Nano* **4**, 7048–7052 (2010).

- Jung, S., Tiwari, M. K., Doan, N. V. & Poulikakos, D. Mechanism of supercooled droplet freezing on surfaces. *Nature Commun.* **3**, 615 (2012).
- Chandra, S. & Avedisian, C. T. On the collision of a droplet with a solid surface. *Proc. R. Soc. Lond. A* **432**, 13–41 (1991).
- Clanet, C., Beguin, C., Richard, D. & Quéré, D. Maximal deformation of an impacting drop. *J. Fluid Mech.* **517**, 199–208 (2004).
- Rein, M. Phenomena of liquid-drop impact on solid and liquid surfaces. *Fluid Dyn. Res.* **12**, 61–93 (1993).
- Eggers, J., Fontelos, M. A., Josserand, C. & Zaleski, S. Drop dynamics after impact on a solid wall: theory and simulations. *Phys. Fluids* **22**, 062101 (2010).
- Richard, D. & Quéré, D. Bouncing water drops. *Europhys. Lett.* **50**, 769–775 (2000).
- Bartolo, D. *et al.* Bouncing or sticky droplets: impalement transitions on superhydrophobic micropatterned surfaces. *Europhys. Lett.* **74**, 299–305 (2006).
- Deng, T. *et al.* Nonwetting of impinging droplets on textured surfaces. *Appl. Phys. Lett.* **94**, 133109 (2009).
- Kwon, Y., Patankar, N., Choi, J. & Lee, J. Design of surface hierarchy for extreme hydrophobicity. *Langmuir* **25**, 6129–6136 (2009).
- Bergeron, V., Bonn, D., Martin, J. Y. & Vovelle, L. Controlling droplet deposition with polymer additives. *Nature* **405**, 772–775 (2000).
- Bartolo, D., Josserand, C. & Bonn, D. Retraction dynamics of aqueous drops upon impact on non-wetting surfaces. *J. Fluid Mech.* **545**, 329–338 (2005).
- Reyssat, M., Richard, D., Clanet, C. & Quéré, D. Dynamical superhydrophobicity. *Faraday Discuss.* **146**, 19–33 (2010).
- Li, X. Y., Ma, X. H. & Lan, Z. Dynamic behavior of the water droplet impact on a textured hydrophobic/superhydrophobic surface: the effect of the remaining liquid film arising on the pillars' tops on the contact time. *Langmuir* **26**, 4831–4838 (2010).
- Bird, R. B., Stewart, W. E. & Lightfoot, E. N. *Transport Phenomena* (Wiley, 1960).
- Okumura, K., Chevy, F., Richard, D., Quéré, D. & Clanet, C. Water spring: a model for bouncing drops. *Europhys. Lett.* **62**, 237–243 (2003).
- Taylor, G. The dynamics of thin sheets of fluid. III. Disintegration of fluid sheets. *Proc. R. Soc. Lond. A* **253**, 313–321 (1959).
- Culick, F. E. C. Comments on a ruptured soap film. *J. Appl. Phys.* **31**, 1128–1129 (1960).
- Rayleigh, L. On the capillary phenomena of jets. *Proc. R. Soc. Lond.* **29**, 71–97 (1879).
- Wachters, L. H. J. & Westerling, N. A. J. The heat transfer from a hot wall to impinging water drops in the spheroidal state. *Chem. Eng. Sci.* **21**, 1047–1056 (1966).
- Corrigan, R. D. & DeMiglio, R. D. *Effect of Precipitation on Wind Turbine Performance*. Report No. NASA-TM-86986 (NASA, 1985).
- Cotton, K. C. *Evaluating and Improving Steam Turbine Performance* (Cotton Fact, 1993).

Supplementary Information is available in the online version of the paper.

Acknowledgements K.K.V. was supported by a DARPA Young Faculty Award, the MIT Energy Initiative, an NSF Career Award (0952564) and the MIT-Deshpande Center. J.C.B. was supported by an NSF Postdoctoral Research Fellowship (DMS1004678). We thank T. Buonassisi, Y. Cui, A. Paxson and J. Bales at MIT for the use of equipment, and D. Quéré, G. McKinley, J. Bush, K. Corriveau and P. Barbone for reading and commenting on the manuscript.

Author Contributions J.C.B., R.D., H.-M.K. and K.K.V. designed the research; J.C.B., R.D. and H.-M.K. performed the research; J.C.B., R.D., H.-M.K. and K.K.V. analysed the data; J.C.B. wrote the original manuscript and all authors helped revise it. K.K.V. supervised the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.K.V. (varanasi@mit.edu).

METHODS

Drop bouncing and imaging. Simultaneous top and side views of drop impact were captured with two high-speed cameras, each filming at a minimum of 10,000 frames per second. A combination of high-speed cameras (Photron S1, Phantom v7 and Colour Phantom v5) were used in these experiments. The drops were released from a needle at a fixed height above the surface. The size of the drop, the impact velocity and the contact time were calculated directly from the high-speed images for each trial.

Laser-ablated silicon surfaces. Control surfaces were fabricated by irradiating silicon surfaces with 100-ns pulses at a repetition rate of 20 kHz from an Nd:YAG laser at 1,064 nm wavelength and 150 W maximum continuous output. The surface was kept normal to the direction of the incident beam. Desired patterns were produced by rastering the laser beam with multiple steps. After coating with trichloro(1H,1H,2H,2H-perfluorooctyl)silane, the surface became superhydrophobic with an advancing contact angle of $\sim 163^\circ$ and a receding contact angle of $\sim 161^\circ$. These surfaces (control) display minimal pinning, as indicated by the extremely low contact angle hysteresis, $\sim 2^\circ$.

The ridge surface was designed such that the height varies as $z = a \sin^n(x/\lambda)$, where x is the horizontal distance and a , n and λ are constant parameters. The values of these parameters were selected as $\lambda = 4$ mm (to allow the drop to interact with one or two peaks regardless of impact locations), $a = 150$ μm (to provide a feature amplitude large enough to influence the film thickness h) and $n = 100$ (to restrict the full-width at half-maximum of the texture to 300 μm , a value small enough not to significantly influence the film thickness h away from the peak).

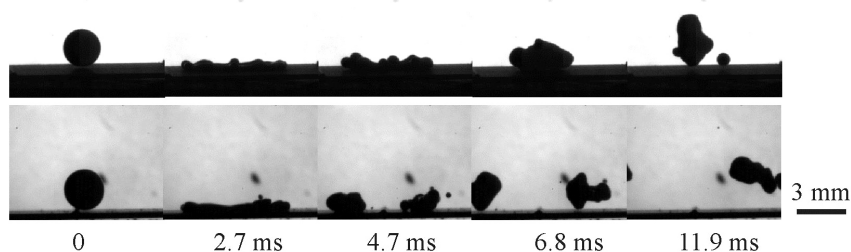
Silicon micropillar surface. The silicon micropillar array used in the experiments was fabricated using standard photolithography processes. A photomask with square windows was used and the pattern was transferred to photoresist using ultraviolet light exposure. Next, reactive ion etching in inductively coupled plasma was used to etch the exposed areas to form micropillars (each micropillar was 10 μm square with 10 μm height and was separated from the next pillar by 5 μm). Trichloro(1H,1H,2H,2H-perfluorooctyl)silane was coated onto the micropillars using vapour-phase deposition to render the surface superhydrophobic (advancing contact angle $\sim 165^\circ$, receding contact angle $\sim 132^\circ$).

Anodized aluminium oxide surface. The anodized aluminium oxide (AAO) surface was prepared by a two-step anodization and etching process. A 40 mm \times 40 mm square and 5 mm thick piece of aluminium (grade 6061) was milled in a CNC machine to have ridges of 100 μm height and 200 μm width, as shown in Extended Data Fig. 4a. The surface was then thoroughly cleaned by first sonicating in acetone followed by rinsing with ethanol and distilled water and drying with nitrogen. The surface was first electropolished with a mixture of perchloric acid and ethanol (in a ratio of 1:3, respectively) for 20 min at 20 V and 100 mA. During this process, the mixture was stirred and maintained at 7°C with the help of a stirrer plate. The surface was then washed several times with distilled water and then dried using nitrogen. After electropolishing, the surface was anodized with phosphoric acid for one hour at 40 V while the acid was continuously stirred and

maintained at 15°C . The surface was again thoroughly washed with distilled water and dried with nitrogen. The surface was then ready for etching, which was done with a mixture of chromic and phosphoric acids that were dissolved in distilled water in a proportion of 1.6 wt% and 6 wt%, respectively. The etching was done for 45 min while the mixture was maintained at 65°C and continuously stirred. After this step, the surface was thoroughly washed with distilled water, dried with nitrogen and kept overnight in a refrigerator. The etching step was repeated at the same conditions for 2 h. Finally, the surface was cleaned thoroughly with distilled water and dried with nitrogen. SEM images (Extended Data Fig. 4b, c) of the anodized surface reveal that it has a hierarchical structure consisting of micropits (~ 10 –50 μm) and nanometre-scale pores (~ 50 –100 nm). A drop of water placed on the surface spread completely, indicating that the surface was superhydrophilic. To render the surface hydrophobic, it was coated with trichloro(1H,1H,2H,2H-perfluorooctyl)silane using vapour-phase deposition. To characterize the hydrophobicity, contact angles were measured with a goniometer and found to be about 159° (advancing) and 157° (receding), indicating the surface was superhydrophobic with minimal pinning.

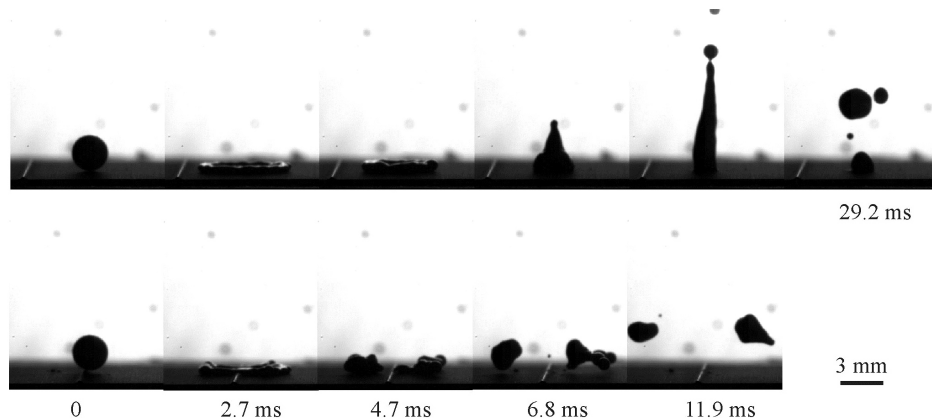
Copper oxide surface. The 100 μm high and 200 μm wide ridges were milled on a copper block, as for the AAO surface. Then, the following steps³¹ were carried out to fabricate spiky nanostructures on the surface. The milled copper plate was ultrasonically cleaned in 3 M hydrochloric acid for 10 min, and rinsed with deionized water. Then, the plate was treated in a 30 mM sodium hydroxide solution, kept at 60°C , for 20 h, followed by multiple rinses with deionized water and drying with nitrogen. The treated surface shows spike-like nano-scale textures, shown in Extended Data Fig. 5. Then, the surface was coated with trichloro(1H,1H,2H,2H-perfluorooctyl)silane using vapour-phase deposition to render it superhydrophobic.

1. Liu, J. *et al.* Hierarchical nanostructures of cupric oxide on a copper substrate: controllable morphology and wettability. *J. Mater. Chem.* **16**, 4427–4434 (2006).
2. Aziz, S. D. & Chandra, S. Impact, recoil, and splashing of molten metal droplets. *Int. J. Heat Mass Transfer* **43**, 2841–2857 (2000).
3. Legendre, D., Daniel, C. & Guiraud, P. Experimental study of a drop bouncing on a wall in a liquid. *Phys. Fluids* **17**, 097105 (2005).
4. Reyssat, M., Pépin, A., Marty, F., Chen, Y. & Quéré, D. Bouncing transitions on microtextured materials. *Europhys. Lett.* **74**, 306 (2006).
5. Jung, Y. C. & Bhushan, B. Dynamic effects of bouncing water droplets on superhydrophobic surfaces. *Langmuir* **24**, 6262–6269 (2008).
6. Brunet, P., Lapiere, F., Thomy, V., Coffinier, Y. & Boukherroub, R. Extreme resistance of superhydrophobic surfaces to impalement: reversible electrowetting related to the impacting/bouncing drop test. *Langmuir* **24**, 11203–11208 (2008).
7. Tuteja, A., Choi, W., Mabry, J., McKinley, G. H. & Cohen, R. E. Robust omniphobic surfaces. *Proc. Natl Acad. Sci. USA* **105**, 18200–18205 (2008).
8. Tsai, P., Pacheco, S., Pirat, C., Lefferts, L. & Lohse, D. Drop impact upon micro- and nanostructured superhydrophobic surfaces. *Langmuir* **25**, 12293–12298 (2009).
9. Zou, J., Wang, P. F., Zhang, T. R., Fu, X. & Ruan, X. Experimental study of a drop bouncing on a liquid surface. *Phys. Fluids* **23**, 044101 (2011).
10. Kwon, D. H. & Lee, S. J. Impact and wetting behaviors of impinging microdroplets on superhydrophobic textured surfaces. *Appl. Phys. Lett.* **100**, 171601 (2012).



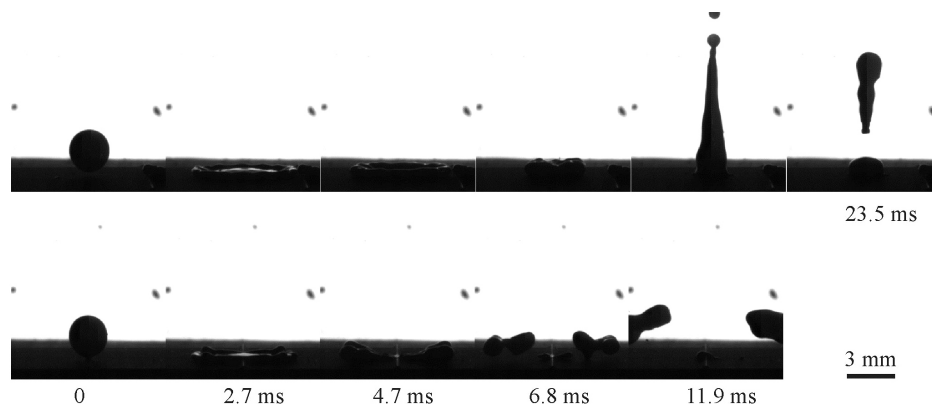
Extended Data Figure 1 | Impact of molten tin droplets (250 °C) on microscopically textured silicon substrates without (top row) and with (bottom row) macroscopic ridges. The substrate temperature is 150 °C, 82 °C

below the droplet freezing point. In both cases, the droplets are able to bounce off the substrate.



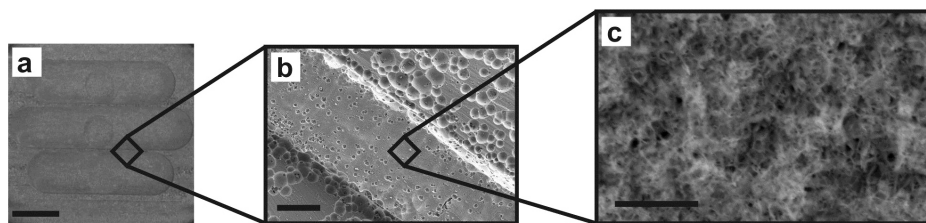
Extended Data Figure 2 | Impact of molten tin droplets (250 °C) on microscopically textured substrates without contacting (top row) and contacting (bottom row) a macroscopic ridge. Here the substrate is maintained at 125 °C (a subcooling of 107 °C). When the droplet hits the

macroscopic ridge, it is able to bounce off in 6.8 ms, whereas when impact is not on the ridge, the droplet is arrested owing to solidification. For more details, see Supplementary Video 5.



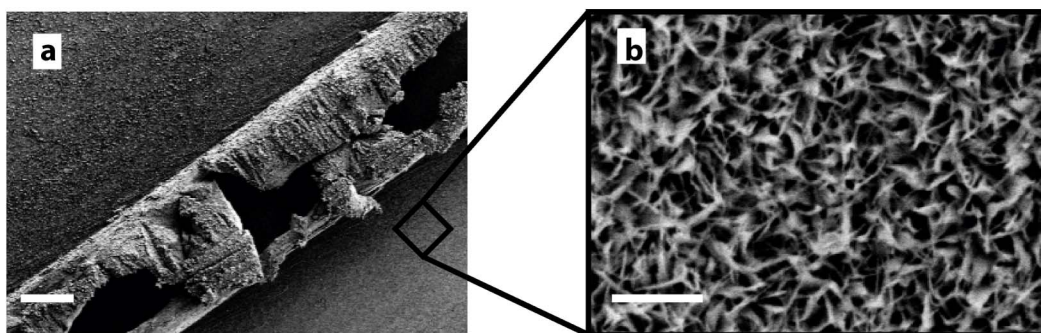
Extended Data Figure 3 | Impact of molten tin droplets (250 °C) on microscopically textured silicon substrates without (top row) and with (bottom row) ridges. Droplets impacting the ridge surface continued to bounce off until the substrate was cooled to about 50 °C, indicating that a

significantly large subcooling (~ 182 °C) is needed to arrest the droplets on the ridge surface. Droplets impacting the surface without ridges (maintained at 50 °C) is arrested owing to solidification.



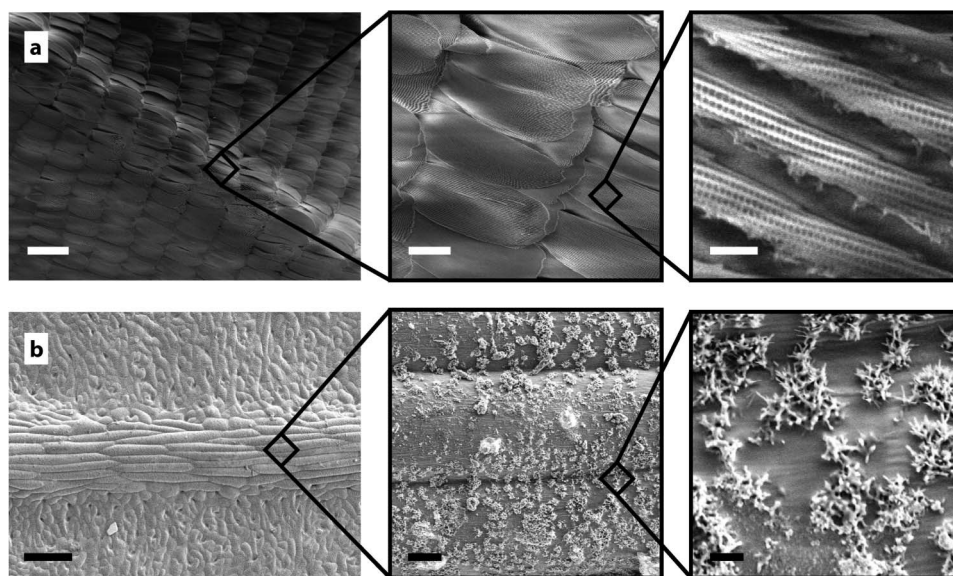
Extended Data Figure 4 | Images of AAO substrate surface at different magnifications. **a**, Top view of the anodized aluminium oxide (AAO) surface showing the macro-scale ridges (height $\sim 100\ \mu\text{m}$, width $\sim 200\ \mu\text{m}$); scale bar,

5 mm. **b**, Magnified SEM image of a single ridge showing micropits; scale bar, 100 μm . **c**, Further magnified SEM image showing nanoscale pores; scale bar, 1 μm .



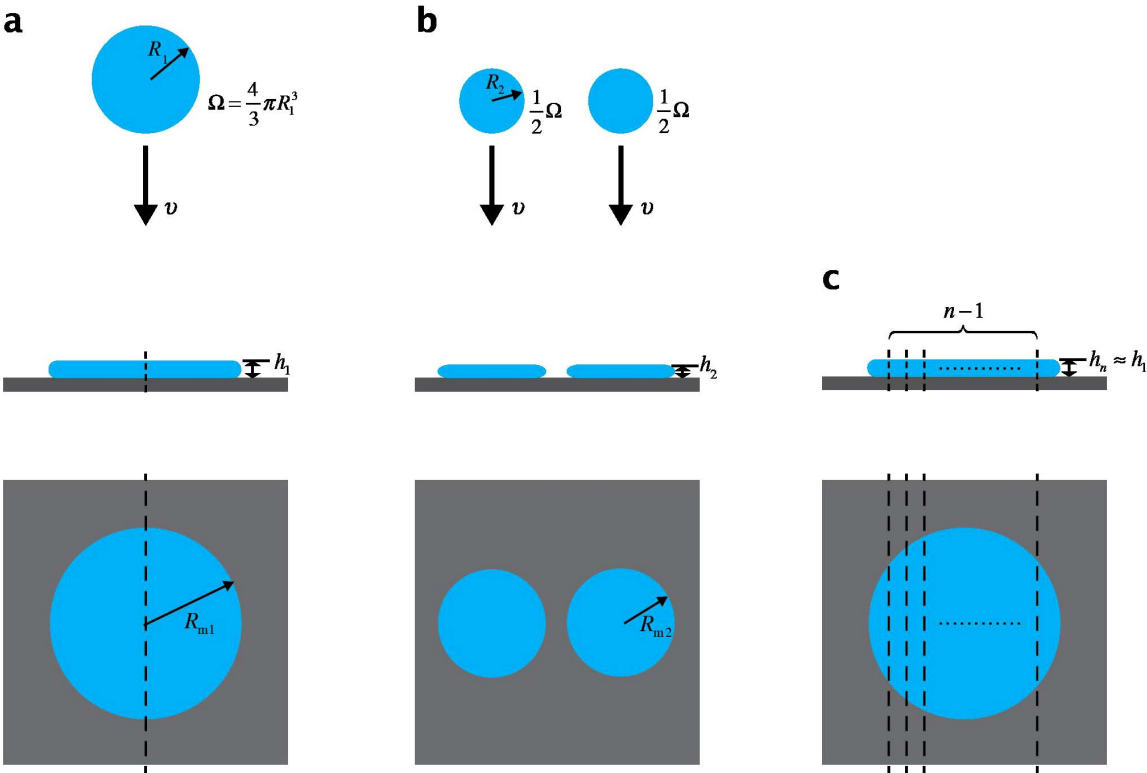
Extended Data Figure 5 | Images of copper oxide substrate surface at different magnifications. **a**, SEM image of the copper oxide nano-textured

macro-ridge (height $\sim 100\ \mu\text{m}$, width $\sim 200\ \mu\text{m}$); scale bar, 100 μm . **b**, A magnified image, showing spiky nano-textures; scale bar, 1 μm .



Extended Data Figure 6 | SEM images of naturally occurring surfaces at different magnifications. **a**, A vein on the wing of a Morpho butterfly

(*M. didius*); **b**, a vein on a nasturtium leaf (*T. majus* L.). Scale bars in **a** left to right; 200 μm , 50 μm and 1 μm ; scale bars in **b** left to right; 200 μm , 10 μm and 2 μm .



Extended Data Figure 7 | Droplet splitting and contact time. a–c, Diagrams of the ridge case (a), the simplistic case where a droplet splits before impact (b), and the generalized (*n*-split parts) ridge case (c).

Extended Data Table 1 | Experimental contact time of bouncing drops from past studies

Study	Droplet	Radius (mm)	Contact time (ms)	Contact time (dimensionless)	Ref.
Wachters & Westerling (1966)	water on hot solid	1.15	11.1	2.4	28
Richard & Quéré (2000)	water	0.4	2.6	3	15
Aziz & Chandra (2000)	molten tin	1.35	13	2.3	32
Richard et al. (2002)	water	0.1-5	0.3-50	2.6	1
Clanet et al. (2004)	water	1.25	13.5	2.6	12
Bartolo et al. (2005)	water	1	16	4	20
Legendre et al (2005)	toluene in water	1.3	28	3.0	33
Bartolo et al. (2006)	water	1	15	4	16
Reyssat et al. (2007)	water	1.2	13 ± 2	3	34
Jung & Bhushan (2008)	water	1	16	4	35
Brunet <i>et al</i> (2008)	water	1.35	23	4.0	36
Tuteja et al (2008)	hexadecane	0.72	350	110	37
Tsai et al. (2009)	water	1	12.5	3	38
Reyssat et al. (2010)	water	1.15	13	2.8	21
Mishchenko et al. (2010)	water	1.5	20	2.9	8
Li et al. (2010)	water	1.35	14.9-22.3	2.5-3.8	22
Zou et al. (2011)	water on water	0.86-2.33	15-62	4.8 [§]	39
Kwon & Lee (2012)	water	0.022	0.032	2.6	40
This paper	water	1.3	7.8	1.4	
	§ Liquid impact on liquid. All other studies liquid drop impacts solid				

Experimental evidence for the influence of group size on cultural complexity

Maxime Derex¹, Marie-Pauline Beugin¹, Bernard Godelle¹ & Michel Raymond^{1,2}

The remarkable ecological and demographic success of humanity is largely attributed to our capacity for cumulative culture^{1–3}. The accumulation of beneficial cultural innovations across generations is puzzling because transmission events are generally imperfect, although there is large variance in fidelity. Events of perfect cultural transmission and innovations should be more frequent in a large population⁴. As a consequence, a large population size may be a prerequisite for the evolution of cultural complexity^{4,5}, although anthropological studies have produced mixed results^{6–9} and empirical evidence is lacking¹⁰. Here we use a dual-task computer game to show that cultural evolution strongly depends on population size, as players in larger groups maintained higher cultural complexity. We found that when group size increases, cultural knowledge is less deteriorated, improvements to existing cultural traits are more frequent, and cultural trait diversity is maintained more often. Our results demonstrate how changes in group size can generate both adaptive cultural evolution and maladaptive losses of culturally acquired skills. As humans live in habitats for which they are ill-suited without specific cultural adaptations^{11,12}, it suggests that, in our evolutionary past, group-size reduction may have exposed human societies to significant risks, including societal collapse¹³.

The accumulation of socially learned information over many generations has enabled humans to develop powerful technologies that no individual could have invented alone¹⁴. Cumulative culture is most likely to be restricted to the *Homo* genus and remains an evolutionary puzzle¹⁵. Several hypotheses have been proposed to explain this explosion in cultural complexity, with a recent emphasis on social-learning mechanisms specific to humans, such as teaching, language or imitation^{16,17}. These mechanisms of faithful transmission stabilize cultural knowledge, thus enabling successive improvements, as has been previously shown theoretically¹⁸ and empirically^{19,20}. However, perfect transmission is most probably unrealistic, as for any given transmission event, an information loss is expected, particularly for complex tasks^{4,21}. Moreover, transmission is only one aspect of the problem, as cumulative cultural evolution also requires the creation of new knowledge; that is, innovation.

The determinants of technological regression—the opposite situation—have been studied in Tasmanian aboriginals. It was argued that cultural losses were associated with population-size reduction²². A general model of cultural evolution that links demographic factors to psychological aspects of social learning has been proposed by Henrich⁴. Considering that transmission events for complex tasks are generally imperfect, with a large variance in fidelity, a learner could acquire by chance greater skill than the demonstrator if the number of transmission events (that is, the population size) is sufficiently large. As there is a psychological propensity to imitate successful individuals (prestige bias), this individual becomes the new demonstrator, driving cultural evolution. A decrease in population size makes such events unlikely, making cultural regression unavoidable. Analytical modelling shows that, as the population size increases, the combination of imperfect learning and prestige bias can lead to cumulative evolution, even if transmission is generally inaccurate. Bursts of cultural complexity during

the Palaeolithic era (2.6 million years ago to 10 thousand years ago) and particularly during the Upper Palaeolithic transition (45 thousand years ago) may illustrate demographic processes, rather than changes in cognitive abilities^{5,23}. However, factors favouring the ability to develop complex culture will most probably also have a positive effect on population size, thus limiting causal assessments using correlative studies. Furthermore, studies using anthropological data produced mixed results^{6–9}. The only experimental study to investigate how group size influences cumulative cultural evolution reported no relationship¹⁰. However, only one cultural task was considered, and the larger group size was limited to three individuals. More parameters must be explored experimentally to investigate the effect of group size on cultural complexity.

Following Henrich's analysis, the maintenance of a cultural task within a group should depend on group size and task complexity. Specifically, within a group of a particular size, greater loss of information is expected for a more complex task. Alternatively, for a task of a given complexity, greater loss of information is expected in a smaller group. Thus, when considering two improvable tasks, one simple and one complex, artificially introduced into groups of different sizes, we predict that the simple task will be better conserved than the complex task (prediction 1); the probability of conserving the complex task will increase with group size (prediction 2); and better performance will be observed in the larger groups for both tasks (prediction 3).

To study the effect of group size on cultural complexity, 366 men participated in a dual-task computer game. Players had to collect resources individually to improve their 'health'. A cultural package composed of two demonstrations, one concerning a simple task and one concerning a complex task, was introduced within groups of different sizes (2, 4, 8 or 16 players). The players were told that each item in the cultural package could be improved. During each of the 15 trials of the game, each player had to build an arrowhead (simple task) or a fishing net (complex task) to collect 'life units' (see Extended Data Fig. 1). The cultural trait diversity of the group thus consisted of some players building one artefact, while the remaining players built the other; diversity was lost when all individuals built the same object.

As expected from prediction 1, the simple task was more likely than the complex task to be maintained for all group sizes ($\chi^2 = 3.83$, d.f. = 1, $P = 0.05$; Fig. 1a, b). For each task, the probability of being lost (none of the individuals of the group exploited the task at the end of the game, see Methods) by a group decreased with increasing group size ($\chi^2 = 7.62$, d.f. = 1, $P = 0.006$), as expected from prediction 2 (Fig. 1a, b). Interestingly, the increased probability of maintaining the complex task in large groups did not reduce the probability of maintaining the simple task (type of task \times group size interaction $\chi^2 = 0.85$, d.f. = 1, $P = 0.36$). Indeed, the probability of maintaining cultural diversity (that is, observing both tasks in the group) increased with group size ($\chi^2 = 16.3$, d.f. = 1, $P < 0.0001$; Fig. 1c).

For each group size, the performances of the best within-group artefacts (simple and complex) at the fifteenth trial were compared to the score of the equivalent artefact from the cultural package. The simple task was stable in the smaller groups and improved in the larger groups (Fig. 2). A linear model was used to investigate the effect of

¹University of Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. ²CNRS, Institute of Evolutionary Sciences, CC 065, Place Eugène Bataillon, Montpellier, France.

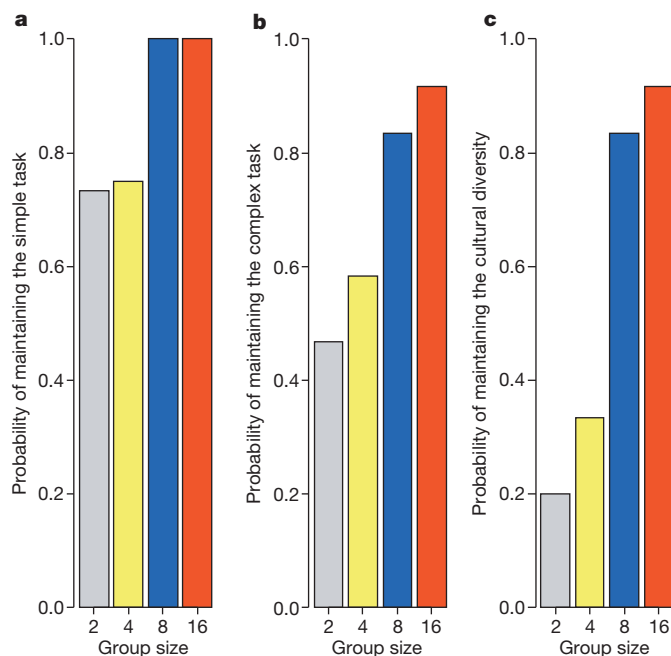


Figure 1 | Group size affects the maintenance of cultural tasks.

a–c, Probability of at least one observation of the simple task (**a**), the complex task (**b**) or both (that is, cultural diversity) (**c**) among the three last trials, for group size of 2 ($n = 15$ replicates), 4 ($n = 12$), 8 ($n = 12$) and 16 ($n = 12$) players.

group size and shows that group size had a linear effect on the performance of the best within-group arrowhead, suggesting that cultural evolution was enhanced in larger groups, consistent with prediction 3 ($F_{1,48} = 10.2$, $P = 0.003$; Fig. 2 and Extended Data Fig. 2). Performance of the complex task deteriorated in the smaller groups and remained stable in the larger groups (Fig. 3). Group size had a linear and quadratic effect on the performance of the best within-group fishing net ($F_{1,47} = 7.12$, $P = 0.01$ and $F_{1,47} = 4.22$, $P = 0.05$, respectively; Fig. 3). Among groups maintaining the complex task, only the 8- and 16-player groups improved it compared to the original cultural package (see Extended Data Figs 3 and 4).

The improvement of both tasks was linked to group size, suggesting that refinement of pre-existing technology is facilitated by increasing group size. The link between innovation rate and group size is not

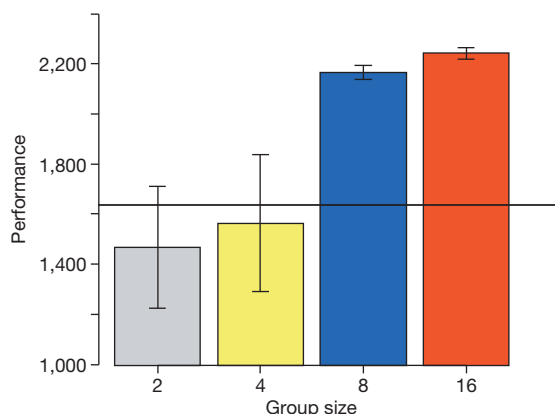


Figure 2 | Larger groups favour improvements to the simple cultural trait.

The horizontal line shows the arrowhead performance from the cultural package. Performance is measured using arbitrary life units. Plotted are the mean values \pm s.e.m. The simple task was stable in the smaller groups (mean performance: 2-player groups = 1,466, $t = -0.71$, d.f. = 14, $P = 0.49$; 4-player groups = 1,563, $t = -0.27$, d.f. = 11, $P = 0.79$) and improved in the larger groups (8-player groups = 2,166, $t = 18.84$, d.f. = 11, $P < 0.0001$; 16-player groups = 2,242, $t = 27.57$, d.f. = 11, $P < 0.0001$).

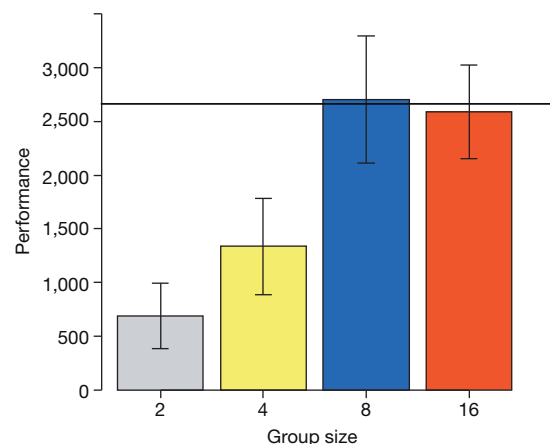


Figure 3 | Larger groups prevent degradation of the complex cultural trait.

The horizontal line shows the fishing-net performance from the cultural package. Performance is measured using arbitrary life units. Plotted are the mean values \pm s.e.m. The complex task deteriorated in the smaller groups (mean performance: 2-player groups = 685, $t = -6.50$, d.f. = 14, $P < 0.0001$; 4-player groups = 1,334, $t = -2.99$, d.f. = 11, $P = 0.01$) and remained stable in the larger groups (mean performance: 8-player groups = 2,706, $t = 0.07$, d.f. = 11, $P = 0.95$; 16-player groups = 2,590, $t = -0.17$, d.f. = 11, $P = 0.87$).

surprising, as the combination of inter-individual variance in cognitive abilities and sampling effect increase the probability of observing high performers within a large group. Furthermore, a group can collectively achieve a solution to a cognitive problem that is not available to an individual through ‘swarm intelligence’²⁴. Whatever the mechanism, the best within-group artefacts drove the performance of the entire group, as shown by the correlation between best within-group artefacts and other within-group artefacts at the final trial (arrowhead, Pearson correlation = 0.39, $t = 5.53$, d.f. = 167, $P < 0.0001$; fishing net, Pearson correlation = 0.29, $t = 2.78$, d.f. = 87, $P = 0.007$).

When technological complexity is measured by the number of existing tools in the cultural repertoire, archaeological data produce mixed results^{6–9}. The occurrence of new tools is poorly understood, but individuals rarely invent new tools from scratch; pre-existing technologies should have a role through combination; that is, bringing together two established cultural traits to generate a new trait^{18,25,26}. Interestingly, this game suggests that increasing group size favours the maintenance of cultural diversity, a prerequisite for subsequent innovation through combination. It is worthy of note that the aim of the game was to maximize the player’s ‘health’. Thus, a player not able to perform the complex task (for example, lacking good visual memory) could perform better by efficiently repeating the simple task than by trying the complex one. It suggests that the individual diversity associated with larger group size could be pivotal to the maintenance of cultural trait diversity. By facilitating the maintenance of cultural diversity, increasing group size could also favour the emergence of division of labour at the group level. Such conditions pave the way for the emergence of inter-individual collaborations and group-level organization, some of the most important properties of human groups²⁷.

At the individual level, results also show that complex-task (fishing net) copying was most of time associated with a loss of skill, whereas simple-task copying was not (see Supplementary Information). This confirms that greater loss of information is expected for a more complex task, as suggested by Henrich⁴. At the group level, the maintenance of the complex task observed in large groups is thus explained by an increased probability to observe rare events directly linked to group size, such as a perfect copy or even an innovation, rather than overall better individual copying abilities. Following an innovation, prestige bias leads individuals to shift, and copy a new model. Even if copying deteriorates information, the mean group performance can increase, allowing cultural evolution to operate⁴. Accordingly, cultural

complexity—as measured in the archaeological record, for example—is most probably not a direct marker of the mean cognitive ability, as an ecological increase in population size could trigger the onset of a cumulative cultural evolution. Such an event may subsequently lead to the evolution of advanced copying ability, as this trait will most probably be an advantage in such a cultural environment. The players' difficulty in properly copying the fishing net from the cultural package (100% of fishing-net builders failed at the first trial) also illustrates the importance of multiple demonstrations and multiple attempts in the acquisition process²⁸. In our game, players acquired the correct skill over several trials. In large groups, high-performing copiers (more likely to be observed as group size increases) can prevent the skill from disappearing, enabling players who lack good copying ability to benefit from more demonstrations.

Our results support Henrich's hypothesis: changes in group size can generate both adaptive cultural evolution and maladaptive losses of culturally acquired skills⁴. In our evolutionary past, group-size reduction may have exposed human societies to notable risks, as humans live in many habitats to which they are ill-suited without specific cultural adaptations^{11,12}. Indeed, the more that we depend for our survival on large bodies of culturally transmitted knowledge, the more we rely on living in large groups. Under such conditions, group-size reduction could have triggered important loss of skills, leading to societal collapse¹³, particularly in challenging environments. Interestingly, some cumulative cultural innovations, such as writing, printing and various forms of long-term data storage, allow the preservation of information outside of individuals, such that it is unknown whether the maintenance of current cultural complexity is nowadays similarly dependent on group size.

METHODS SUMMARY

Each player was randomly assigned to a group of 2, 4, 8 or 16 players, and all groups started the game by benefiting from the same cultural package (composed of an arrowhead and a fishing net, see Methods section for the complete details of the game). The simple task involved drawing an arrowhead, for which the performance evaluation depended only on its shape. The arrowhead demonstration in the cultural package involved 15 steps and provided 1,638 life units. The complex task involved building a fishing net, for which the performance evaluation depended on its shape and the procedure used to build it. The fishing-net demonstration in the cultural package involved 39 steps (the sequence of which mattered) and provided 2,665 life units. The starting individual life level was 3,400 units, and 1,000 units (daily needs) were subtracted at each trial. The task difficulties were designed so that, for a non-experienced player, the probability of scoring below their daily needs (and thus having a negative score) was low when choosing the arrowhead task and high when choosing the fishing-net task. Each trial was followed by an information period during which players could choose a single demonstration to observe (ranked by their performance), from one of their group members or the cultural package. The cultural package was available up to the third trial: from the fourth trial and after, social information came only from players' group members. A total of 366 male students (mean age = 24.1 years, s.d. = 4.4) played this game only once, in groups of 2 (15 replicates), 4, 8 or 16 (12 replicates each) players.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 August; accepted 14 October 2013.

Published online 13 November 2013.

1. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (Univ. of Chicago Press, 1985).

2. Tomasello, M., Kruger, A. C. & Ratner, H. H. Cultural learning. *Behav. Brain Sci.* **16**, 495–511 (1993).
3. Boyd, R. & Richerson, P. J. *Not by Genes Alone* (Univ. of Chicago Press, 2005).
4. Henrich, J. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses: the Tasmanian case. *Am. Antiq.* **69**, 197–214 (2004).
5. Powell, A., Shennan, S. & Thomas, M. G. Late Pleistocene demography and the appearance of modern human behavior. *Science* **324**, 1298–1301 (2009).
6. Kline, M. A. & Boyd, R. Population size predicts technological complexity in Oceania. *Proc. R. Soc. B* **277**, 2559–2564 (2010).
7. Read, D. Population size does not predict artifact complexity: analysis of data from tasmania, arctic hunter-gatherers, and oceania fishing groups. *UC Los Angeles: Hum. Complex Systems* <http://www.escholarship.org/uc/item/61n4303q> (2012).
8. Collard, M., Kemery, M. & Banks, S. Causes of toolkit variation among hunter-gatherers: a test of four competing hypotheses. *Can. J. Archaeol.* **29**, 1–19 (2005).
9. Read, D. An interaction model for resource implement complexity based on risk and number of annual moves. *Am. Antiq.* **73**, 599–625 (2008).
10. Caldwell, C. A. & Millen, A. E. Human cumulative culture in the laboratory: effects of (micro) population size. *Learn. Behav.* **38**, 310–318 (2010).
11. Henrich, J. & Broesch, J. On the nature of cultural transmission networks: evidence from Fijian villages for adaptive learning biases. *Phil. Trans. R. Soc. Lond. B* **366**, 1139–1148 (2011).
12. Boyd, R., Richerson, P. J. & Henrich, J. The cultural niche: why social learning is essential for human adaptation. *Proc. Natl Acad. Sci. USA* **108**, 10918–10925 (2011).
13. Diamond, J. The Tasmanians — the longest isolation, the simplest technology. *Nature* **273**, 185–186 (1978).
14. Boyd, R. & Richerson, P. J. Why does culture increase human adaptability. *Ethol. Sociobiol.* **16**, 125–143 (1995).
15. Boyd, R. & Richerson, P. J. Why culture is common, but cultural evolution is rare? *Proc. Br. Acad.* **88**, 77–93 (1996).
16. Tomasello, M. In *Social Learning in Animals: the Roots of Culture* (eds Heyes, C. M. & Galef, B. G.) 319–346 (Academic, 1996).
17. Tennie, C., Call, J. & Tomasello, M. Ratcheting up the ratchet: on the evolution of cumulative culture. *Phil. Trans. R. Soc. Lond. B* **364**, 2405–2415 (2009).
18. Lewis, H. M. & Laland, K. N. Transmission fidelity is the key to the build-up of cumulative culture. *Phil. Trans. R. Soc. Lond. B* **367**, 2171–2180 (2012).
19. Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B. & Laland, K. N. Identification of the social and cognitive processes underlying human cumulative culture. *Science* **335**, 1114–1118 (2012).
20. Derex, M., Godelle, B. & Raymond, M. Social learners require process information to outperform individual learners. *Evolution* **67**, 688–697 (2013).
21. Claidière, N. & Sperber, D. Imitation explains the propagation, not the stability of animal culture. *Proc. R. Soc. B* **277**, 651–659 (2010).
22. Diamond, J. *Guns, Germs, and Steel: the Fates of Human Societies* (W. W. Norton, 1999).
23. Marquet, P. A. et al. Emergence of social complexity among coastal hunter-gatherers in the Atacama Desert of northern Chile. *Proc. Natl Acad. Sci. USA* **109**, 14754–14760 (2012).
24. Krause, J., Ruxton, G. D. & Krause, S. Swarm intelligence in animals and humans. *Trends Ecol. Evol.* **25**, 28–34 (2010).
25. Basalla, G. *The Evolution of Technology* (Cambridge Univ. Press, 1988).
26. Henrich, J. & Boyd, R. On modeling cognition and culture: Why cultural evolution does not require replication of representations. *J. Cogn. Cult.* **2**, 87–112 (2002).
27. Smaldino, P. E. The cultural evolution of emergent group-level traits. *Behav. Brain Sci.* (in the press).
28. Flynn, E. & Whiten, A. Studying children's social learning experimentally “in the wild”. *Learn. Behav.* **38**, 284–296 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Belkhir for help in establishing the Structured Query Language database, D. Dubois for recruiting participants and organizing the experimental sessions, and the Laboratory of Experimental Economics of Montpellier (University of Montpellier I) for hosting the experiment. Contribution ISEM 2013-146.

Author Contributions M.D., B.G. and M.R. designed the study. M.-P.B. and M.D. collected data. M.D., M.P.B. and M.R. analysed the data. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D. (maxime.derex@gmail.com).

METHODS

Participants. A total of 366 male students were randomly selected from a database managed by the Laboratory of Experimental Economics of Montpellier (LEEM) and recruited by email from various universities in Montpellier (Southern France). The subjects ranged in age from 18 to 49 years (mean = 24.1 years, *s.d.* = 4.4 years). Each participant was randomly assigned to one condition of the experiment. Participants received fees for travel according to the LEEM operating rule (€2 for local participants, €6 for others).

Procedure. The experiment took place in a computer room at the LEEM. For each session, a maximum of 20 players sat at physically separated and networked computers and were randomly assigned to a group (the number of players per group varied according to the treatment, see below). They could not see each other, and they were blind with regard to the purpose of the experiment and who belonged to each group. The players were instructed that communication was not allowed. The participants could read instructions on their screens about the rewards and the goal of the game, and they were requested to enter their date of birth before the start of the game. At the end of the game, each subject received a reward according to his performance (€10 on average, see rewards calculation).

Principle. The participants played a computer game (programmed in Object Pascal with Delphi 6) during which they had to maximize their 'health' using two virtual tasks, making an arrowhead or a fishing net. Before the beginning of the game, players were advised that the fishing-net task was potentially more effective than the arrowhead task but that the fishing-net construction was more difficult. The participants were also informed that the performance of an arrowhead depended only on its shape, whereas the performance of the fishing net depended on its shape and the procedure used to build it. Each player began the game by observing a video demonstration of each task from a cultural package and was instructed that the arrowhead and fishing-net demonstrations could be improved. The arrowhead demonstration involved 15 steps and was associated with a score of 1,638. The fishing-net demonstration involved 39 steps (the sequence of which mattered) and was associated with a score of 2,665. The participants were not aware of the highest achievable score for any task.

The players then had 15 trials to collect resources and improve their health score. At each trial, they had the opportunity to build either an arrowhead or fishing net. Players began the game with a health score of 3,400 units. At each trial, their health level was reduced by 1,000 units, corresponding to their daily needs. Between trials, players could benefit from social information (see below).

Construction period. During the construction period (limited to 90 s), the players had to choose between the arrowhead task and the fishing-net task to collect resources.

The arrowhead task. The performance of an arrowhead depended only on its shape. The arrowhead score ranged from 0 to 2,400 units. A simple symmetric, triangular arrowhead constituted an acceptable performance equal to the player's daily needs. As a consequence, the probability of a non-experienced player scoring below his daily needs was low.

Construction details for the arrowhead task. First, the players had to choose the rectangular grid dimension on which to draw the arrowhead (30 possible values, Extended Data Fig. 1.a). Once the grid was chosen, the players had to draw their arrowhead. By clicking on the grid, the players could draw lines between points (Extended Data Fig. 1.b). The players had to draw the outline of their arrowhead and the virtual relief. No construction rules were implemented.

Score calculation for the arrowhead task. Once an arrowhead was drawn, it was evaluated by the program. The arrowhead was scanned pixel by pixel to evaluate five parameters: the size (α) and the symmetry (β) of the arrowhead, the number of notches (γ) and their regularities (δ), and the triangular shape (λ). All the parameters were compared to a theoretical optimal value and normalized from 0 to 1. The score S was then obtained according to this formula:

$$S = \alpha.400 + \beta.400 + \gamma.800 + \delta.400 + \lambda.400 \quad (1)$$

The fishing-net task. The participants had access to several virtual tools with which to build their nets. The performance of a net depended on its shape and the procedure used to build it. The net's score ranged from 0 to 5,135 units. Departure from the construction rules (which were unknown to the players) resulted in increased penalties during use of the fishing net. As a consequence, the probability of a non-experienced player scoring below his daily needs (1,000 units) was high.

Construction details for the fishing-net task. First, the players had to choose the squared grid dimension on which to build the net (30 possible values, Extended Data Fig. 1.c). Once the grid was chosen, the players had access to different types of ropes and knots, as in a previous experiment²⁰. A rope could be set between any pair of attaching points, and a knot could be tied to any attaching point, in any order (Extended Data Fig. 1.d). There were limited ropes and knots available. Each

additional rope placed on the frame decreased the length of the remaining rope according to the length used. This remaining quantity was visible on the screen. There were three different types of rope available (thick (red), medium (blue) and thin (green)). Each additional knot placed on the net decreased the length of the remaining knot quantity according to the type of knot used (three sizes available). This remaining quantity of knots was visible on the screen. Modification of one parameter produced complex interactions with others to generate a complex fitness landscape. For example, the use of the thickest ropes prevented the net from breaking but increased the net visibility so that the number of potentially caught fish was reduced. In addition, the order of construction (the process), was important. For example, two ropes that intersect at an attaching point should be tied together with a knot before another rope is put on the frame. If this step is omitted, the expected score is reduced.

Score calculation for the fishing-net task. Once a fishing net was constructed, it was evaluated by the program. A global resistance score (GR) was calculated according to the number of knots and compared to the required number. A local resistance score (LR_{*i*}) was determined for each mesh *i* according to the length and thickness of the ropes involved. During each virtual fishing exercise, 79 fish were launched, with a unique size of 65 (arbitrary units). The probability of each fish encountering the net increased according to the net overall size (set by the grid-spacing parameter) and decreased according to its visibility. The visibility of a net was computed as the sum of the lengths of all ropes used, weighted by their thicknesses. Once a fish was set to interact with the net, random coordinates were generated to identify at which mesh the interaction took place. If the fish was smaller than the mesh, it escaped. If it was larger, the probability of the net breaking was calculated as $1 - (GR \cdot LR_i)$. In such a case, the whole fishing process stopped. If the net did not break, the fish could escape with a probability P_{esc} , which depended on the shape of the mesh and construction-rule penalty. If the fish did not escape, its size was added to the player's score. This process was repeated until the last fish was encountered or until the net broke.

Information period. After each trial, the resulting score, along with the player's health level, was displayed. The players could also see score lists for the arrowheads and fishing nets generated by the player's group members at the previous trial, ordered by performance. During the first three trials, the cultural package (arrowhead or fishing net) was included in the corresponding list.

By clicking on a score, the players could see the step-by-step procedure needed to build the selected item. Any demonstration lasted 40 s, regardless of the number of building steps. At each information period, a player could see only one demonstration. From the fourth information period, cultural-package demonstrations were removed from the lists. The players then had access only to their group member's demonstrations. The duration of the social-information period was 70 s.

Rewards calculation. The individual rewards were €10 on average. Players who died during the game (health level dropped below 0) earned €2. The other players earned an amount €A calculated according to this formula:

$$A = H_p/H_g \cdot [5 \cdot N + 3 \cdot N_d] + 5 \quad (2)$$

where H_p is the player's health level, H_g is the sum of the group's health levels, N is the size of the group, and N_d is the number of dead players within the group.

Treatments. Four group sizes were considered: 2 players, 4 players, 8 players and 16 players. All treatments were replicated 12 times, except for the 2-player treatment, which was replicated 15 times.

Cultural evolution. The aim of the study was to investigate the evolution of the cultural packages that were introduced in the experimental groups. Two types of analyses were carried out; one examined the maintenance of cultural tasks (whether some individuals exploited the cultural task at the end of the game), and the other examined the performance associated with the tasks. For each of the two tasks, we focused on the best within-group information because this information drives subsequent cultural evolution (due to prestige bias).

Maintenance of cultural tasks. Two models were used. One model investigated how the simple task was maintained in comparison with the complex task. A cultural task was considered to be maintained within a group if, among the last three trials, at least one individual of the group exploited the task. The response variable was the presence or absence of each task in each group. The independent variables were the type of task (arrowhead or fishing net), group size, mean age within the group, and type of task \times group size interaction, with 'group identity' as a random factor. Generalized linear mixed models (binomial) were used.

The other model investigated how cultural diversity was maintained according to group size. Cultural diversity was considered to be observed within a group if, in the last three trials, at least one individual performed the arrowhead task, whereas at least one individual performed the fishing-net task. The response variable was the presence or absence of the diversity. The independent variables were the group size and mean age within the group. A generalized linear model (binomial) was used.

Best within-group information. The performances of the best within-group arrowheads at the fifteenth trial were compared to the score of the arrowhead from the cultural package, using a one-sample Student's *t*-test (if the distribution significantly departed from normality, a Mann–Whitney–Wilcoxon test was also performed; results were qualitatively similar, data not shown). A further linear model was used to investigate the effect of group size. In this case, the response variable was the score of the best within-group arrowhead at the fifteenth trial, and the independent variables were group size and mean age within the group. These two analyses were carried out again for the fishing-net performances.

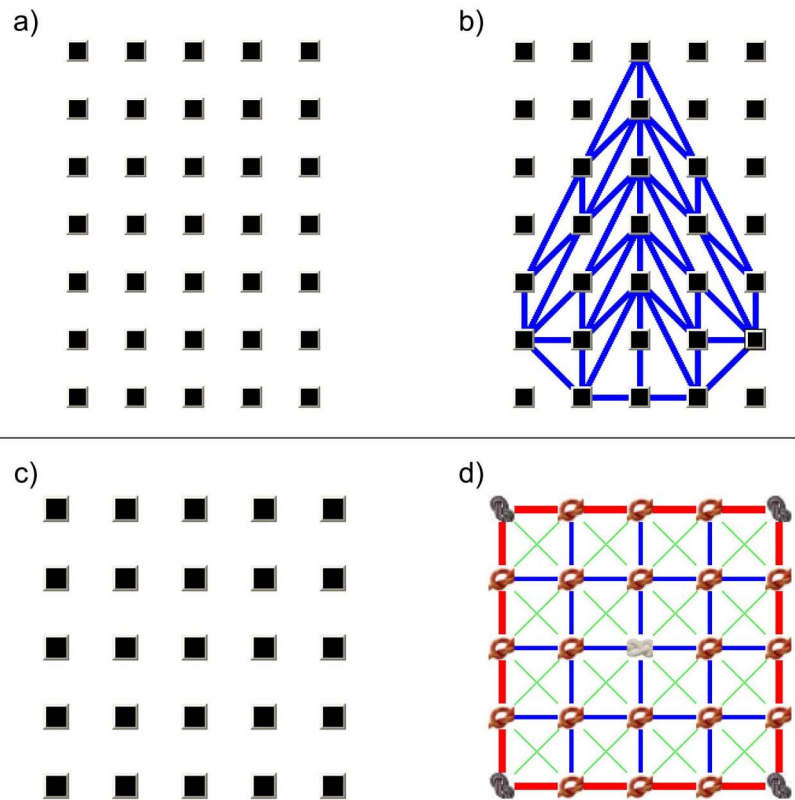
As groups could lose one of the two tasks, all analyses were carried out twice. In one case, we considered all groups, and performance score of zero was assigned when a task was lost from a group, that is, the degradation of the performance was considered complete (results shown in the main text). In the other case, we considered only the performance of the groups that conserved the task (results shown in Extended Data Figs 2 and 3).

Normality of residuals was significantly rejected (using Shapiro's test) in three models. This was owing to the presence of zero values (associated with task loss) generating a gap in the distribution between zero and the minimal score. When the presence or absence of the task was explicitly controlled for in order to estimate this gap, normality of residuals were not rejected (sometimes requiring the exclusion of only one outlier). All results described here were unchanged, whether or not these changes were made.

Fidelity of copying. Henrich's model assumes that information transmission is generally imperfect (particularly with complex tasks). Indeed, if copying is faithful, no cultural losses are expected. For each task, analyses were carried out to evaluate copying fidelity. During the observation period, players could choose a single demonstration to observe before building a new artefact. The aim was to study whether or not artefacts built by the players performed worse than the artefacts they observed.

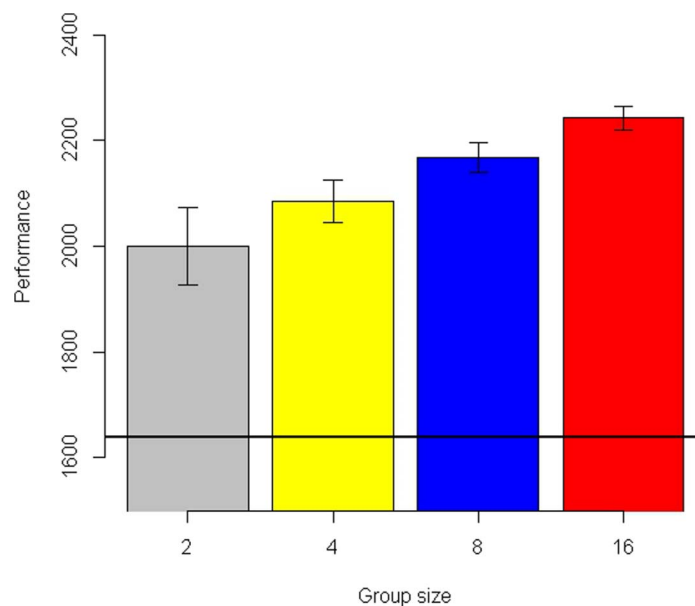
An observed artefact was considered as a model and was associated with *n* copies, depending on how many players observed the same model. For example, if three players observed the same model, three copies (copy 1, copy 2 and copy 3) were created. All possible pairs of artefacts were formed from the model and the copies: with one model and three copies, this corresponded to 6 pairs (model–copy 1; model–copy 2; model–copy 3; copy 1–copy 2; copy 1–copy 3; copy 2–copy 3). Comparisons of 'model–copy' represent our treatment of interest: if copying deteriorates information, the expected score difference (model score minus copy score) should be positive (null or negative otherwise). Comparisons of 'copy–copy' represent a control treatment: the expected score difference should be null. The focal artefact (first artefact from the pair) was either a copy or the model and was always compared to a copy (second artefact from the pair). The skill was considered to have deteriorated when the focal artefact outperformed the copy (score difference strictly positive). The binary response variable was the presence or absence of skill degradation. The independent variables were the type of the focal artefact ('copy' or 'model'). The identity of the focal artefact and the identity of the producer of the second artefact from the pair were included as random effects. A generalized linear mixed model (binomial) was used. All analyses were carried out separately for each task (arrowhead and fishing net).

Correlation between best within-group information and individual performances. This study was culture-centred, focusing on the state of the information available within groups (how the best within-group information performed). Considering that the best-within-group information influences the subsequent performance of the entire group, it is important to test the correlation between best within-group information and individual performances: owing to prestige bias, the best within-group information should affect the performance of the entire group. We examined the correlation between the best within-group information and the performance of the other players at the fifteenth trial using the Pearson correlation test.



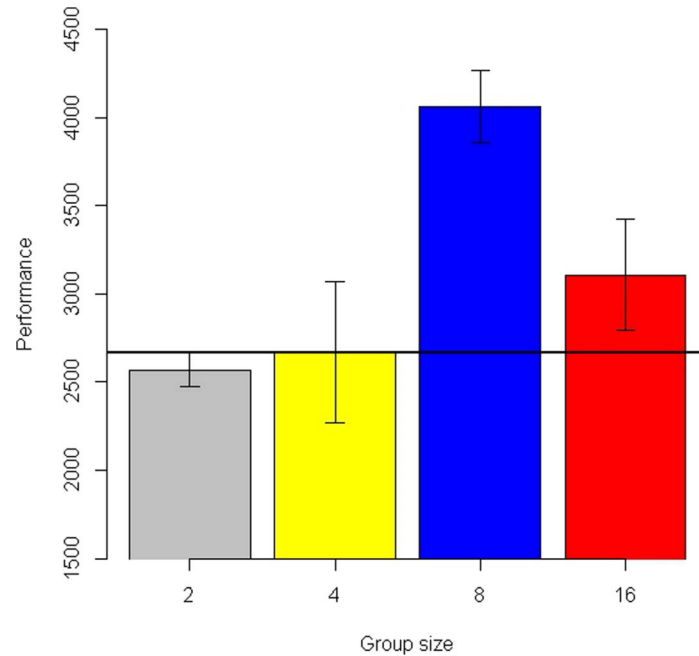
Extended Data Figure 1 | Cultural tasks. **a**, Rectangular grid composed of 35 attaching points in which to draw an arrowhead. The spacing between the attaching points was modifiable. **b**, An example of an arrowhead. **c**, Square grid

composed of 25 attaching points in which to build a fishing net. The spacing between the attaching points was modifiable. **d**, An example of a fishing net.



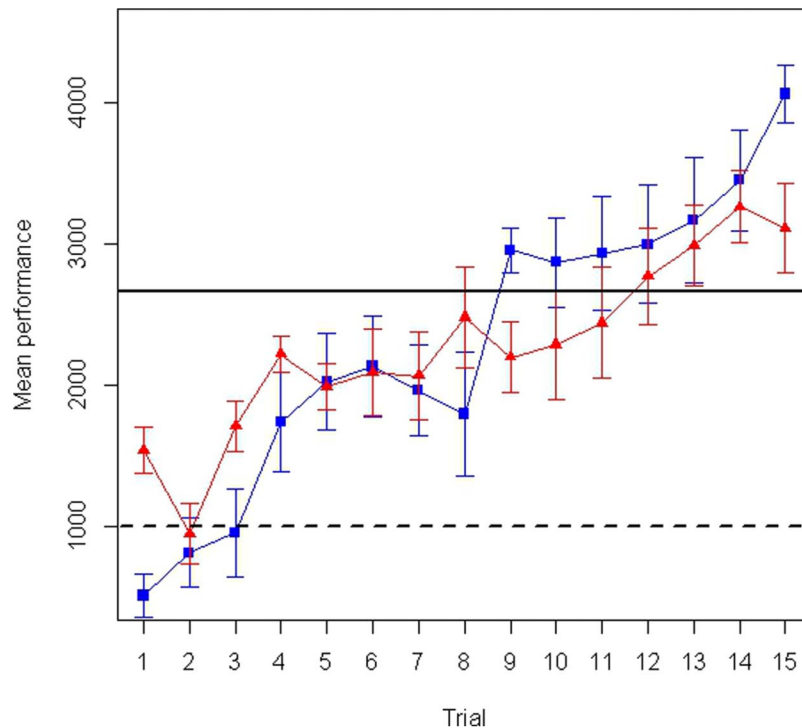
Extended Data Figure 2 | Best within-group information associated with the simple task, when conserved within the group. Performance is measured using arbitrary life units. Plotted are the mean values \pm s.e.m. Considering only the performance of the groups that conserved the task (see Methods), the simple task of the cultural package was improved in all group sizes (mean performance: 2-player groups = 2,000, $t = 4.90$, d.f. = 10, $P = 0.0006$; 4-player

groups = 2,085, $t = 11.12$, d.f. = 8, $P < 0.0001$; 8-player groups = 2,166, $t = 18.84$, d.f. = 11, $P < 0.0001$; 16-player groups = 2,242, $t = 27.57$, d.f. = 11, $P < 0.0001$). Group size had a linear effect on the performance of the best within-group arrowhead ($F_{1,41} = 15.3$, $P = 0.0003$). The horizontal line shows the performance of the arrowhead from the cultural package.



Extended Data Figure 3 | Best within-group information associated with the complex task, when conserved within the group. Performance is measured using arbitrary life units. Plotted are the mean values \pm s.e.m. Only 4 2-player groups (26.7%) conserved the complex task and were therefore excluded from the analysis. The complex task was stable in the 4-player groups (mean performance = 2,669, $t = 0.01$, d.f. = 5, $P = 0.99$) and improved in the larger groups. The difference between 8-player groups and the demonstration of the cultural package was significant (mean = 4,059, $t = 6.79$, d.f. = 7, $P = 0.0001$, one-sided) but marginally significant concerning 16-player groups (mean = 3,108, $t = 1.40$, d.f. = 9, $P = 0.09$, one-sided). Group size had a linear

and an unexpected quadratic effect on the performance of the best within-group fishing net ($F_{1,24} = 10.6$, $P = 0.003$ and $F_{1,24} = 9.88$, $P = 0.004$, respectively). This quadratic effect could indicate that participants had trouble making use of the information in a large group, but our experimental design allows us to rule out this possibility (see Supplementary Information). Instead, early performances of 16-player groups affected the probability of observing the cultural-package demonstration, hindering players from acquiring pivotal information (see Extended Data Fig. 4 and Supplementary Information). The horizontal line shows the performance of the fishing net from the cultural package.



Extended Data Figure 4 | Best within-group information associated with a fishing net (when conserved within the group) across time. The red line shows 16-player groups and the blue line shows 8-player groups. Performance is measured using arbitrary life units. Plotted are the mean values \pm s.e.m. At the beginning of the game, the 16-player groups performed better than the 8-player groups ($F_{1,22} = 21.7$, $P = 0.0001$), as expected. However, the opposite was observed at the end of the game ($F_{1,16} = 5.68$, $P = 0.03$). During the first three trials, the performance associated with the best within-group fishing net affected the probability of observing the cultural-package demonstration. Thus, the probability of observing the cultural-package demonstration was lower in

16-player groups compared with 8-player groups. A lower rate of observation of the cultural-package reduced the group performance suggesting that the observation of demonstrations from other sources hindered the acquisition of pivotal information (see Supplementary Information for details). It suggests that, under specific conditions, the increasing number of valuable sources of information associated with larger group size could lead to a suboptimal cultural evolution rate. The horizontal solid line shows the performance of the fishing net from the cultural package. The horizontal dashed line shows the players' daily needs.

A canonical to non-canonical Wnt signalling switch in haematopoietic stem-cell ageing

Maria Carolina Florian¹, Kalpana J. Nattamai², Karin Dörr¹, Gina Marka¹, Bettina Überle¹, Virag Vas¹, Christina Eckl³, Immanuel Andrä⁴, Matthias Schiemann⁴, Robert A. J. Oostendorp³, Karin Scharffetter-Kochanek¹, Hans Armin Kestler⁵, Yi Zheng² & Hartmut Geiger^{1,2}

Many organs with a high cell turnover (for example, skin, intestine and blood) are composed of short-lived cells that require continuous replenishment by somatic stem cells^{1,2}. Ageing results in the inability of these tissues to maintain homeostasis and it is believed that somatic stem-cell ageing is one underlying cause of tissue attrition with age or age-related diseases. Ageing of haematopoietic stem cells (HSCs) is associated with impaired haematopoiesis in the elderly^{3–6}. Despite a large amount of data describing the decline of HSC function on ageing, the molecular mechanisms of this process remain largely unknown, which precludes rational approaches to attenuate stem-cell ageing. Here we report an unexpected shift from canonical to non-canonical Wnt signalling in mice due to elevated expression of Wnt5a in aged HSCs, which causes stem-cell ageing. Wnt5a treatment of young HSCs induces ageing-associated stem-cell apolarity, reduction of regenerative capacity and an ageing-like myeloid–lymphoid differentiation skewing via activation of the small Rho GTPase Cdc42. Conversely, Wnt5a haploinsufficiency attenuates HSC ageing, whereas stem-cell-intrinsic reduction of Wnt5a expression results in functionally rejuvenated aged HSCs. Our data demonstrate a critical role for stem-cell-intrinsic non-canonical Wnt5a signalling in HSC ageing.

Aged muscle stem cells can regenerate muscles as efficiently as young muscle stem cells either by forced activation of Notch, or by parabiosis-mediated inhibition of Wnt signalling^{7–10}. Whether there is a similar critical role of Wnt signalling in ageing of HSCs remains largely unexplored. Multiple members of the Wnt family are expressed in haematopoietic cells as well as in non-haematopoietic stroma cells (for a concise review see ref. 11). Wnt3a (associated with canonical Wnt signalling) and Wnt5a (associated with non-canonical signalling) are so far the most studied Wnt proteins in haematopoiesis^{12–16}.

Notably, high levels of Wnt5a as well as Wnt4 mRNA were detected in middle-aged (10 months) and aged (20–24 months old) long-term (LT)-HSCs (Lin[−], Sca-1⁺, c-Kit⁺, CD34[−], Flk2[−]) and Lin[−] cells from C57BL/6 as well as DBA/2 mice, whereas they were almost absent in young (2–3 months old) cells (Fig. 1a and Extended Data Fig. 1a–c), concurrently with elevated Wnt5a protein levels in aged haematopoietic cells (Fig. 1b, c). Other Wnt proteins (including canonical-signalling-associated Wnt1, Wnt3a, Wnt5b and Wnt10b) did not present with changes in expression on ageing. In young LT-HSCs, Wnt5a localizes mainly at the plasma membrane, whereas aged LT-HSCs showed Wnt5a distributed primarily within the cytoplasm (Extended Data Fig. 1d–g and Supplementary Video 1). Wnt5a localization only partially overlapped with clathrin-positive vesicular structures in aged LT-HSCs (Extended Data Fig. 1f).

In young LT-HSCs, β -catenin is localized mainly in the nucleus, indicative of active canonical Wnt signalling (Fig. 1d, e and Supplementary Video 2). Wnt5a has been reported to directly inhibit canonical Wnt signalling in haematopoietic cells¹². Consistent with this finding, aged

LT-HSCs presented with a reduced level and primarily cytoplasmic localization of β -catenin (Fig. 1d, e and Supplementary Video 3). Reduced levels of β -catenin upon ageing were specific to the LT-HSC compartment, as more differentiated LKs (Lin[−] c-Kit⁺ Sca-1[−] cells), LSKs (Lin[−] Sca-1⁺ c-Kit⁺ cells), lymphoid-primed multipotent progenitors (LMPPs; Lin[−] c-Kit⁺ Sca-1[−] CD34⁺ Flk2⁺ cells) and short-term (ST)-HSCs (Lin[−] c-Kit⁺ Sca-1[−] CD34⁺ Flk2[−] cells) (Extended Data Fig. 1h) showed similar levels of β -catenin upon ageing (Fig. 1g and Extended Data Fig. 1i). Axin2 (an established direct downstream target of canonical Wnt signalling^{9,15}) transcript levels in aged LT-HSCs were markedly decreased (Fig. 1f). Young LT-HSCs treated with Wnt5a elicited a reduction in the level of β -catenin similar to the level found in aged LT-HSCs (Fig. 1g and Extended Data Fig. 1i). The presence of MG-132 (a proteasomal inhibitor) abolishes the reduction of β -catenin, whereas β -catenin degradation is already visible 2 h after Wnt5a exposure (Fig. 1h, i), indicating a direct action of Wnt5a on β -catenin levels. Our data support that on ageing, LT-HSCs shift from canonical to non-canonical Wnt signalling due to, at least in part, elevated Wnt5a expression and signalling in aged LT-HSCs.

In young HSCs a Wnt5a-driven non-canonical signalling pathway regulates quiescence via regulating the activity of the small Rho GTPase Cdc42 (refs 17–19). We recently demonstrated a critical role for elevated Cdc42 activity in ageing and polarity of LT-HSCs²⁰. Bone-marrow-derived haematopoietic progenitor/stem cells treated *in vitro* with Wnt5a showed increased Cdc42 activity (Cdc42–GTP) (Fig. 2a, b). Elevated Cdc42 activity in aged HSCs is associated with a high percentage of LT-HSCs being apolar for tubulin and Cdc42, and this apolarity is a hallmark of aged LT-HSCs^{20,21}. We detected an increased frequency of apolar cells among young Wnt5a-treated LT-HSCs (Fig. 2c, d and Extended Data Fig. 2a) to a level previously described for aged LT-HSCs. When Wnt5a was administered together with casin (a selective inhibitor of Cdc42 activity^{20,22}), the frequency of polarized LT-HSCs did not change (Fig. 2c, d and Extended Data Fig. 2a), supporting the hypothesis that Wnt5a results in apolarity through increasing Cdc42 activity. Notably, Wnt5a treatment also induced apolarity for NCAM2, an adhesion receptor molecule located on the cell membrane (Extended Data Fig. 2b, c). Wnt5a treatment did not alter mRNA expression of Cdc42 (Extended Data Fig. 2d). Rac2 and the Cdc42-related genes *Rhoj* and *Rhoq* showed slightly higher expression in aged compared to young LT-HSCs. Our data support the hypothesis that Wnt5a has a direct effect on Cdc42 activity and polarity establishment, as these changes were induced 2 h after Wnt5a treatment (Fig. 2a, b), although they do not exclude additional indirect effects (see also Extended Data Fig. 5).

Competitive transplant experiments (Extended Data Fig. 2e) revealed that *ex vivo* Wnt5a-treated young LT-HSCs presented with reduced engraftment and an ageing-like skewing in differentiation potential (elevated donor-derived myeloid contribution, reduced donor-derived

¹Department of Dermatology and Allergic Diseases, University of Ulm, 89091 Ulm, Germany. ²Division of Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital Medical Center and University of Cincinnati, Cincinnati, Ohio 45229, USA. ³3rd Department of Internal Medicine, Klinikum rechts der Isar, Technische Universität München, 81675 München, Germany. ⁴Institute of Medical Microbiology, Immunology and Hygiene, Technische Universität München, 81675 München, Germany. ⁵Institut für Neuroinformatik, Department of Bioinformatics and Systems Biology, University of Ulm, 89091 Ulm, Germany.

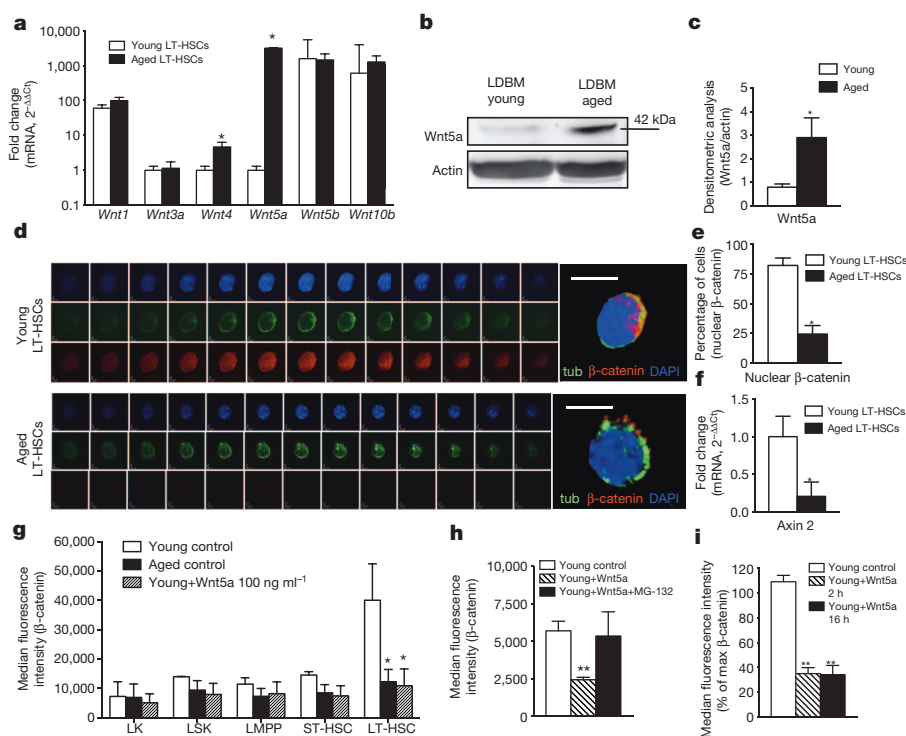


Figure 1 | Increased expression of Wnt5a in aged LT-HSCs results in a shift from canonical to non-canonical Wnt signalling. **a**, Wnt transcript levels in 2–3-month-old or 24-month-old LT-HSCs. $n = 4$, $*P < 0.05$. **b**, **c**, Wnt5a protein levels in low-density bone marrow (LDBM) cells (**b**) and densitometric score (**c**). $n = 4$, $*P < 0.05$. **d**, Immunofluorescence z-stack and three-dimensional merged images of tubulin (green) and β -catenin (red) in LT-HSCs. Scale bar, 5 μ m. **e**, LT-HSCs with nuclear β -catenin. $n = 3$, 200 cells per sample in total. $*P < 0.05$. **f**, Axin 2 transcript levels in LT-HSCs. $n = 4$, $*P < 0.05$. **g**, β -catenin mean fluorescence intensity in young, aged and young

Wnt5a-treated haematopoietic progenitor/stem cells. $n = 3$, $*P < 0.05$ versus young LT-HSC controls. **h**, β -catenin mean fluorescence intensity in young control LT-HSCs, or young LT-HSCs treated with Wnt5a or with Wnt5a plus MG-132. $n = 3$, $**P < 0.01$ versus young controls. **i**, β -catenin mean fluorescence intensity in young LT-HSCs or young LT-HSCs treated with Wnt5a. $n = 3$, $**P < 0.01$ versus young controls. A paired Student's t -test was used to determine the significance of the difference between means of two groups. One-way ANOVA or two-way ANOVA were used to compare means among three or more independent groups. Error bars represent s.e.m.

B-cell contribution; Fig. 2e, f). When Wnt5a treatment was performed in the presence of casin, this ageing-associated differentiation skewing was not observed (Fig. 2f), confirming again the role of a Wnt5a–Cdc42 signalling axis in inducing ageing-like phenotypes in LT-HSCs. Treatment of young LT-HSCs with casin alone had no effect on engraftment potential and differentiation of HSCs, in agreement with previous data²⁰.

We next investigated haematopoiesis in mice haploinsufficient for Wnt5a ($Wnt5a^{+/-}$ mice; $Wnt5a^{-/-}$ mice are embryonic lethal²³). Expression of Wnt5a and the activity of Cdc42 were increased on ageing in wild-type cells (Fig. 3a–d), whereas levels in aged $Wnt5a^{+/-}$ haematopoietic cells were similar to those in $Wnt5a^{+/-}$ and $Wnt5a^{+/+}$ young cells (Fig. 3a–d). Aged $Wnt5a^{+/-}$ LT-HSCs presented with a frequency of

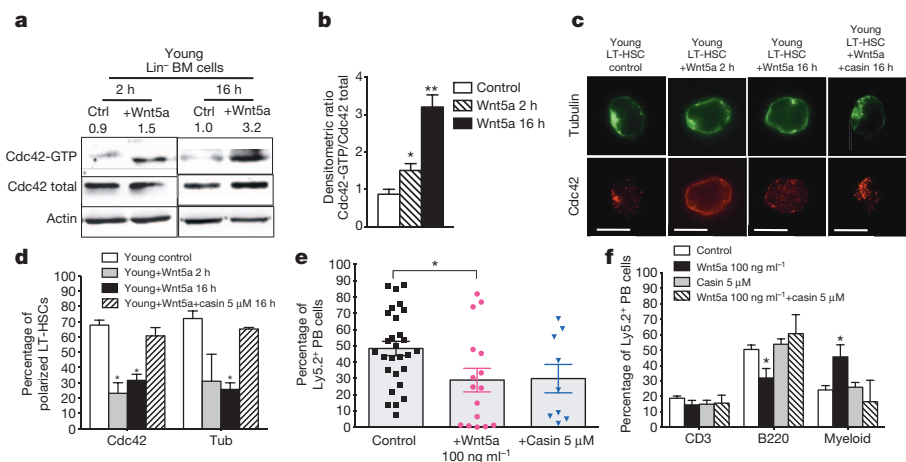


Figure 2 | Wnt5a activates Cdc42 inducing ageing-like phenotypes in young LT-HSCs. **a**, **b**, Cdc42 activity in young and Wnt5a-treated Lin[−] bone marrow (BM) cells (**a**) and densitometric score (**b**). $n = 4$, $*P < 0.05$, $**P < 0.01$. **c**, Cdc42 (red) and tubulin (green) in LT-HSCs shown by immunofluorescence. Scale bar, 5 μ m. **d**, Polar distribution (percentage) of Cdc42 and tubulin in LT-HSCs. $n = 6$, 200 LT-HSCs per sample in total. $*P < 0.001$. **e**, **f**, Donor-derived Ly5.2⁺ cells and B220⁺, CD3⁺ and myeloid (Gr1⁺, Mac1⁺, Gr1⁺Mac1⁺) cells

among Ly5.2⁺ cells in peripheral blood (PB) 24 weeks after transplant. $*P < 0.05$; $n = 10$ for casin and Wnt5a plus casin, $n = 25$ for control and Wnt5a. A paired Student's t -test was used to determine the significance of the difference between means of two groups. One-way ANOVA or two-way ANOVA were used to compare means among three or more independent groups. Error bars represent s.e.m.

polarized cells similar to that found in young *Wnt5a*^{+/-} and *Wnt5a*^{+/+} LT-HSCs (Fig. 3e and Extended Data Fig. 3l, m). The frequency of B cells in aged *Wnt5a*^{+/-} mice compared to aged wild-type mice was significantly increased, paralleled by a decreased myeloid cell frequency (Fig. 3f). Furthermore, total white blood cell, lymphocyte and red blood cell parameters of aged haploinsufficient *Wnt5a*^{+/-} mice were similar to that of young wild-type and young *Wnt5a*^{+/-} mice (Extended Data Fig. 3a–d). Moreover, whereas aged *Wnt5a*^{+/+} mice, as expected, showed an increase in the frequency of both LSKs (Extended Data Fig. 3e) and LT-HSCs (Fig. 3g), aged haploinsufficient *Wnt5a*^{+/-} mice exhibited a reduced frequency of both LSKs and LT-HSCs compared to control aged mice (Fig. 3g and Extended Data Fig. 3e). Notably, expression levels of *Wnt5a* in LT-HSCs, although elevated compared to young LT-HSCs, were almost an order of magnitude lower compared to the level of expression in CD45⁺ stroma cells (Extended Data Fig. 3f). On ageing, levels of *Wnt5a* mRNA in stroma were reduced (Extended Data Fig. 3f). Experiments in which wild-type cells were transplanted into *Wnt5a*^{+/-} and *Wnt5a*^{+/+} recipients revealed that the attenuation of ageing in haematopoiesis in aged *Wnt5a*^{+/-} mice was due to changes in *Wnt5a* levels in haematopoietic cells (Extended Data Fig. 3g–k).

Finally, we asked whether inhibition of LT-HSC cell-intrinsic *Wnt5a* expression through a lentiviral short hairpin RNA (shRNA) approach might functionally rejuvenate aged LT-HSCs in transplantation settings (Fig. 4a and Extended Data Fig. 4a–c). Mice transplanted with aged, untransduced Ly5.2⁺ cells as well as with aged scrambled non-targeting shRNA (NT-GFP⁺Ly5.2⁺) cells showed an elevated frequency of myeloid cells but compromised lymphopoiesis (either B or T cells or both) in both peripheral blood and bone marrow 24 weeks after transplant (Fig. 4b and Extended Data Fig. 4d), and thus the expected aged haematopoietic profile. In contrast, haematopoiesis in mice transplanted with a shRNA specific for *Wnt5a* (*Wnt5a*^{KD}-GFP⁺Ly5.2) presented with improved B lymphopoiesis and a reduction in myeloid skewing, showing a bone marrow and peripheral blood differentiation profile overall

more similar to the one characteristic of young mice (Fig. 4b and Extended Data Fig. 4d). In accordance with stem-cell rejuvenation, the frequency of donor-derived LT-HSCs was reduced in mice transplanted with aged *Wnt5a* knockdown (*Wnt5a*^{KD}) cells compared to mice transplanted with aged non-targeting shRNA or untransduced cells (Fig. 4c). This *Wnt5a* knockdown effect was specific to aged cells because young *Wnt5a* knockdown cells performed very similar to controls (Extended Data Fig. 4e–g). In addition, the activity of Cdc42 was reduced in donor-derived *Wnt5a* knockdown cells whereas total Cdc42 protein levels were unchanged (Fig. 4d, e and Extended Data Fig. 4h), and *Wnt5a*^{KD} LT-HSCs from recipient mice presented with a significantly elevated frequency of polarized cells compared to controls (Fig. 4f and Extended Data Fig. 4i, j). Moreover, aged *Wnt5a*^{KD} LT-HSCs reverted to a high expression level and nuclear localization of β -catenin, and therefore to active canonical Wnt signalling (Fig. 4g, h and Extended Data Fig. 4k). In summary, reduction of ageing-associated elevated *Wnt5a* expression rejuvenates chronologically aged LT-HSCs.

Cross-talk between Wnt and Notch pathways, as reported for muscle stem cells, has not yet been investigated in HSC ageing. Aged LT-HSCs presented with a distinct pattern of expression activation of Notch ligands, receptors and *Hes1* (a direct target indicative of active Notch pathway) (Extended Data Fig. 5a–g). This ageing-related Notch pathway activation was to a great extent recapitulated in young *Wnt5a*-treated LT-HSCs (Extended Data Fig. 5a–g). The non-canonical *Wnt5a* signalling pathway can also trigger Ca²⁺ influx in the cytoplasm and activate CamKII (calmodulin-dependent protein kinase II) and the calcium-sensitive transcription factor NFATc^{23,24}. LT-HSCs presented with a biphasic (a short-term steep increase, followed by a lower, but long-term sustained signal) influx of Ca²⁺ in the cytoplasm in response to *Wnt5a* treatment (Extended Data Fig. 5l) and an increase in the level of p-CamKII (the phosphorylated active fraction of CamKII; total CamKII levels were unaltered or even decreased) and in the level of NFATc as determined by immunofluorescence staining (Extended

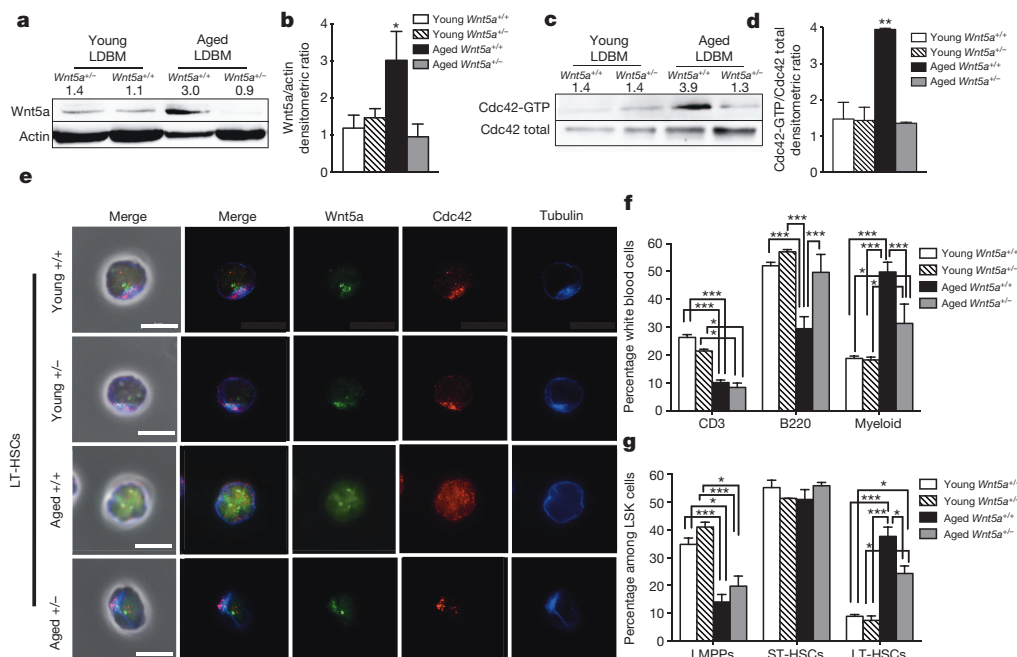


Figure 3 | *Wnt5a* haploinsufficient mice present with attenuated HSC ageing. **a–d**, *Wnt5a* (a) and Cdc42 activity levels (c) in young and aged *Wnt5a*^{+/+} and *Wnt5a*^{+/-} LDBM cells and densitometric analysis (b, d). *n* = 4, **P* < 0.05; ***P* < 0.01. **e**, Distribution of Cdc42, tubulin and *Wnt5a* in young and aged *Wnt5a*^{+/+} and *Wnt5a*^{+/-} LT-HSCs. Scale bar, 5 μ m. **f, g**, Percentage of B220⁺, CD3⁺ and myeloid cells among white blood cells in peripheral blood (f) and percentage of LT-HSCs, ST-HSCs and LMPPs among

LSKs (g) in *Wnt5a*^{+/+} and *Wnt5a*^{+/-} young and aged mice. **P* < 0.05, ***P* < 0.01, ****P* < 0.001; *n* = 5 for aged *Wnt5a*^{+/-} and *Wnt5a*^{+/-} mice; *n* = 7 for young *Wnt5a*^{+/+} and *Wnt5a*^{+/-} mice. A paired Student's *t*-test was used to determine the significance of the difference between means of two groups. One-way ANOVA or two-way ANOVA were used to compare means among three or more independent groups. Error bars represent s.e.m.

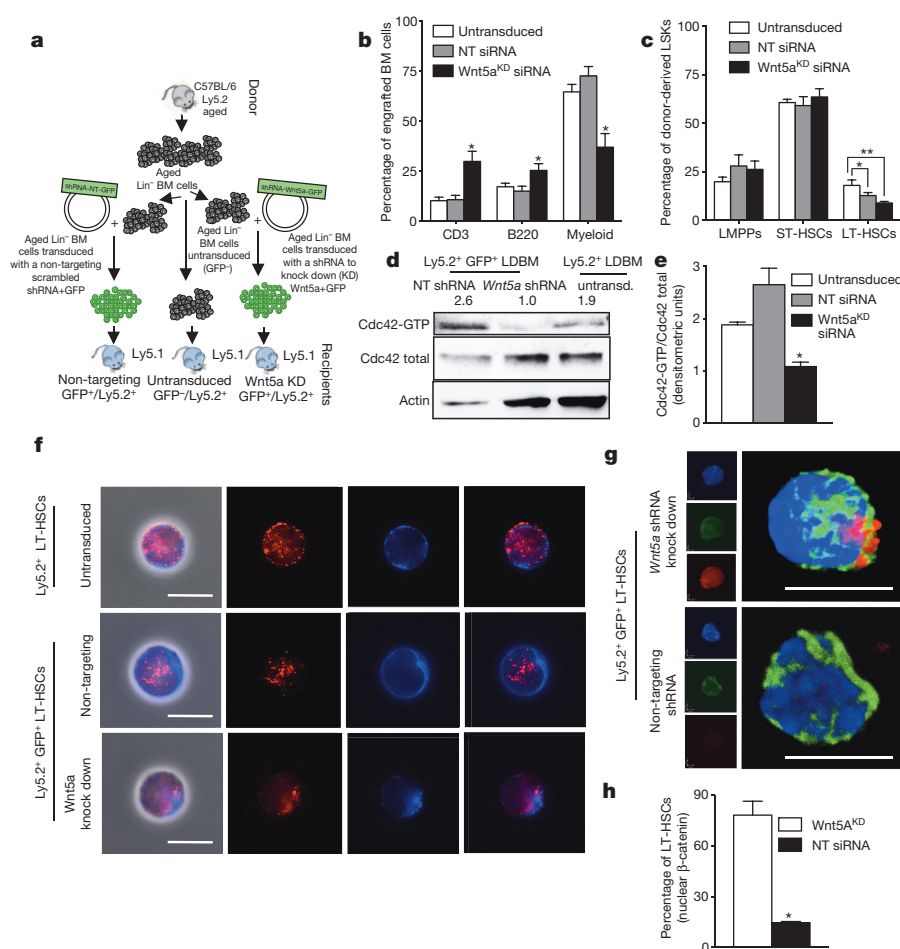


Figure 4 | Reducing Wnt5a expression in aged LT-HSCs rejuvenates their function *in vivo*. **a**, Experimental set-up. **b**, B220⁺, CD3⁺ and myeloid cells among donor-derived cells in bone marrow. **P* < 0.05; *n* = 10. **c**, LT-HSCs, ST-HSCs and LMPPs among donor-derived LSKs. **P* < 0.05; ***P* < 0.01. **d**, **e**, Cdc42 activity in aged donor-derived low-density bone-marrow (LDBM) cells transduced with non-targeting (NT) shRNA, Wnt5a shRNA (Wnt5a^{KD}) and untransduced control (**d**) and densitometric score (**e**). *n* = 4, **P* < 0.05. **f**, Cdc42 and tubulin in donor-derived LT-HSCs from aged untransduced or non-targeting shRNA or Wnt5a^{KD} recipient mice 24 weeks after transplant

Data Fig. 5h–k). Furthermore, the levels of both *p57* and *p27* mRNAs (which have been associated with stem-cell quiescence) were increased in aged and in young Wnt5a-treated LT-HSCs compared to young cells (Extended Data Fig. 5m), whereas cell-cycle parameters of LT-HSCs do not change with age²⁰. These data indicate that there might be cross-talk between the intrinsically increased Wnt5a/Cdc42 non-canonical Wnt pathway and the Notch1 pathway, eliciting Ca²⁺ signalling that ultimately increases quiescence²⁵ and self-renewal of LT-HSCs, which over time translates into an increased number of stem cells with a reduced engraftment potential. Further investigations are warranted to test this hypothesis in more detail (Extended Data Fig. 6).

We present here a novel concept that ageing of LT-HSCs is driven by a shift from canonical Wnt to non-canonical Wnt–Cdc42 signalling. The initiating event is interestingly stem-cell intrinsic (increase of Wnt5a expression in LT-HSCs on ageing). The mechanisms inducing elevated Wnt5a expression in aged LT-HSCs are currently unknown, but possibly involve epigenetic mechanisms²⁶. Our data do not exclude action of stroma-derived Wnt proteins on more differentiated primitive haematopoietic cells or on LT-HSCs in distinct niches^{20,27} or under situations of stress. Whereas we demonstrate a causative role for Wnt5a-induced Cdc42 signalling in HSC ageing, Wnt5a-induced non-canonical signalling could elicit additional changes that might, in addition to Cdc42, contribute to the ageing phenotype.

shown by immunofluorescence. Scale bar, 5 μm. **g**, Immunofluorescence z-stack and three-dimensional merged images of tubulin (green) and β-catenin (red) localization in aged non-targeting shRNA or Wnt5a^{KD} LT-HSCs. Scale bar, 5 μm. **h**, Aged donor-derived non-targeting shRNA or Wnt5a^{KD} LT-HSCs exhibiting nuclear accumulation of β-catenin. *n* = 3, **P* < 0.05. A paired Student's *t*-test was used to determine the significance of the difference between means of two groups. One-way ANOVA or two-way ANOVA were used to compare means among three or more independent groups. Error bars represent s.e.m.

METHODS SUMMARY

C57BL/6 mice (10–12-week-old) were obtained from Janvier. Aged C57BL/6 mice (20–26-month-old) were obtained from the internal divisional stock (derived from mice obtained from both The Jackson Laboratory and Janvier) as well as from NIA/Charles River. Samples were imaged with an AxioObserver Z1 microscope (Zeiss) or with an LSM710 confocal microscope (Zeiss). Primary raw data were imported into the Volocity Software package (Version 6.0, Perkin Elmer) for further processing and conversion into three-dimensional images. For reverse transcriptase real time PCR 20,000–40,000 LT-HSCs from young and aged mice were lysed and processed for RNA extraction immediately after sorting. RNA was obtained with the microRNA extraction kit (Qiagen) and all was used for cDNA conversion. cDNA was prepared and amplified with Ovation RNA Amplification system V2 (NuGEN). Relative levels of GTP-bound Cdc42 were determined by an effector pull-down assay. The lentivirus plasmid vector pLKO.1-YFP and KD sequences were obtained from Sigma's validated genome-wide TRC shRNA libraries.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 July 2012; accepted 3 September 2013.

Published online 20 October 2013.

- Morrison, S. J., Uchida, N. & Weissman, I. L. The biology of hematopoietic stem cells. *Annu. Rev. Cell Dev. Biol.* **11**, 35–71 (1995).
- Fuchs, E. & Segre, J. A. Stem cells: a new lease on life. *Cell* **100**, 143–155 (2000).

3. Rossi, D. J. *et al.* Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc. Natl Acad. Sci. USA* **102**, 9194–9199 (2005).
4. Rossi, D. J., Jamieson, C. H. & Weissman, I. L. Stems cells and the pathways to aging and cancer. *Cell* **132**, 681–696 (2008).
5. Geiger, H. & Van Zant, G. The aging of lympho-hematopoietic stem cells. *Nature Immunol.* **3**, 329–333 (2002).
6. Geiger, H. & Rudolph, K. L. Aging in the lympho-hematopoietic stem cell compartment. *Trends Immunol.* **30**, 360–365 (2009).
7. Conboy, I. M., Conboy, M. J., Smythe, G. M. & Rando, T. A. Notch-mediated restoration of regenerative potential to aged muscle. *Science* **302**, 1575–1577 (2003).
8. Conboy, I. M. *et al.* Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433**, 760–764 (2005).
9. Brack, A. S. *et al.* Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science* **317**, 807–810 (2007).
10. Liu, H. *et al.* Augmented Wnt signaling in a mammalian model of accelerated aging. *Science* **317**, 803–806 (2007).
11. Malhotra, S. & Kincade, P. W. Wnt-related molecules and signaling pathway equilibrium in hematopoiesis. *Cell Stem Cell* **4**, 27–36 (2009).
12. Nemeth, M. J., Topol, L., Anderson, S. M., Yang, Y. & Bodine, D. M. Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation. *Proc. Natl Acad. Sci. USA* **104**, 15436–15441 (2007).
13. Duncan, A. W. *et al.* Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nature Immunol.* **6**, 314–322 (2005).
14. Luis, T. C., Naber, B. A., Fibbe, W. E., van Dongen, J. J. & Staal, F. J. Wnt3a nonredundantly controls hematopoietic stem cell function and its deficiency results in complete absence of canonical Wnt signaling. *Blood* **116**, 496–497 (2010).
15. Luis, T. C. *et al.* Canonical Wnt signaling regulates hematopoiesis in a dosage-dependent fashion. *Cell Stem Cell* **9**, 345–356 (2011).
16. Staal, F. J. & Luis, T. C. Wnt signaling in hematopoiesis: crucial factors for self-renewal, proliferation, and cell fate decisions. *J. Cell. Biochem.* **109**, 844–849 (2010).
17. Schlessinger, K., Hall, A. & Tolwinski, N. Wnt signaling pathways meet Rho GTPases. *Genes Dev.* **23**, 265–277 (2009).
18. Schlessinger, K., McManus, E. J. & Hall, A. Cdc42 and noncanonical Wnt signal transduction pathways cooperate to promote cell polarity. *J. Cell Biol.* **178**, 355–361 (2007).
19. Sugimura, R. *et al.* Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150**, 351–365 (2012).
20. Florian, M. C. *et al.* Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. *Cell Stem Cell* **10**, 520–530 (2012).
21. Geiger, H., de Haan, G. & Florian, M. C. The ageing haematopoietic stem cell compartment. *Nature Rev. Immunol.* **13**, 376–389 (2013).
22. Peterson, J. R., Lebensohn, A. M., Pelish, H. E. & Kirschner, M. W. Biochemical suppression of small-molecule inhibitors: a strategy to identify inhibitor targets and signaling pathway components. *Chem. Biol.* **13**, 443–452 (2006).
23. Liang, H. *et al.* Wnt5a inhibits B cell proliferation and functions as a tumor suppressor in hematopoietic tissue. *Cancer Cell* **4**, 349–360 (2003).
24. Kuhl, M., Sheldahl, L. C., Park, M., Miller, J. R. & Moon, R. T. The Wnt/Ca²⁺ pathway: a new vertebrate Wnt signaling pathway takes shape. *Trends Genet.* **16**, 279–283 (2000).
25. Zou, P. *et al.* p57^{Kip2} and p27^{Kip1} cooperate to maintain hematopoietic stem cell quiescence through interactions with Hsc70. *Cell Stem Cell* **9**, 247–261 (2011).
26. Li, J. *et al.* WNT5A antagonizes WNT/ β -catenin signaling and is frequently silenced by promoter CpG methylation in esophageal squamous cell carcinoma. *Cancer Biol. Ther.* **10**, 617–624 (2010).
27. Kohler, A. *et al.* Altered cellular dynamics and endosteal location of aged early hematopoietic progenitor cells revealed by time-lapse intravital imaging in long bones. *Blood* **114**, 290–298 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Van Zant and J. A. Cancelas for advice and critical reading of the manuscript. We thank F. Kirchhoff and D. van der Merwe for cell sorting support, A. Rück and the Institut für Lasertechnologien in der Medizin und Meßtechnik of Ulm University for support with confocal microscopy, and the Mouse and Cancer Core in Cincinnati and the Tierforschungszentrum of the University of Ulm for supporting our animal work. The work in the laboratory of H.G. is supported by grants from the Deutsche Forschungsgemeinschaft KFO 142, GE2063/1 and SFB 1074, the German Federal Ministry of Education and Research within its joint research project SyStaR (also to H.A.K. and K.S.-K.), the Excellence program of the Baden-Württemberg Foundation, the National Institute of Health, HL076604, DK077762 and AG040118, the Edward P. Evans foundation and the European Commission (FP7 Marie Curie Initial Training Network MARRIAGE). M.C.F. is supported by a 'Bausteinprogramm' of the Department of Medicine of Ulm University.

Author Contributions M.C.F. and H.G. designed and interpreted experiments and wrote the manuscript. M.C.F., K.J.N., B.Ü., C.E., M.S. I.A. and R.A.J.O. performed and analysed experiments. H.G. helped to perform transplantation experiments. V.V. helped with the Wnt5a mouse colony and performed inverse transplantation experiments and prepared sorted stroma CD45⁺ cells. K.D. and G.M. assisted in transplantation procedures and supported in cell sorting and flow analysis procedures. H.A.K. performed expression and database analyses. Y.Z. and K.S.-K. assisted in designing and interpreting experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.G. (hartmut.geiger@uni-ulm.de).

METHODS

Mice. C57BL/6 mice (10–12-week-old) were obtained from Janvier. Aged C57BL/6 mice (20–26-month-old) were obtained from the internal divisional stock (derived from mice obtained from both The Jackson Laboratory and Janvier) as well as from NIA/Charles River. Congenic C57BL/6.SJL-*Ptpr^c*/Boy (BoyJ) mice were obtained from Charles River Laboratories or from the internal divisional stock (derived from mice obtained from Charles River Laboratories). Wnt5a mice were obtained from T. Yamaguchi. All mice were housed in the animal barrier facility under pathogen-free conditions either at the University of Ulm or at CCHMC. All mouse experiments were performed in compliance with the German Law for Welfare of Laboratory Animals and were approved by the Institutional Review Board of the University of Ulm or by the IACUC of CCHMC.

LT-HSC competitive transplantation. For competitive LT-HSC transplantation, young (2–4-month-old) C57BL/6 mice (Ly5.2⁺) were used as donors. Two-hundred LT-HSCs were sorted into 96 multi-well plates and cultured for 16 h in HBSS + 10% FBS with or without 100 ng ml⁻¹ Wnt5a (R&D System) in a water-jacketed incubator at 37 °C, 5% CO₂, 3% O₂. Stem cells were then mixed with 3 × 10⁵ bone marrow cells from young (2–4-month-old) BoyJ competitor mice (Ly5.1⁺) and transplanted into BoyJ recipient mice (Ly5.1⁺). Peripheral blood chimaerism was determined by FACS analysis every 8 weeks up to 24 weeks after primary transplants. The transplantation experiment was performed four times with a cohort of five recipient mice per group each transplant. Inverse transplantation experiments were performed by lethally irradiating recipient *Wnt5a*^{+/+} and *Wnt5a*^{+/-} mice (Ly5.2⁺). 2 × 10⁶ total bone marrow cells from donor BoyJ mice (Ly5.1⁺) were injected into each recipient mouse and transplanted mice were followed and bled every month for up to 20 months after transplantation. In general, transplanted mice were regarded as engrafted when peripheral blood chimaerism was higher or equal to 1.0% and contribution was detected in all lineages.

Flow cytometry and cell sorting. PB and bone marrow cell immunostaining was performed according to standard procedures and samples were analysed on a LSRII flow cytometer (BD Biosciences). Monoclonal antibodies to Ly5.2 (clone 104, eBioscience) and Ly5.1 (clone A20, eBioscience) were used to distinguish donor from recipient and competitor cells. For peripheral blood and bone marrow lineage analysis the antibodies used were all from eBioscience: anti-CD3ε (clone 145-2C11), anti-B220 (clone RA3-6B2), anti-Mac-1 (clone M1/70) and anti-Gr-1 (clone RC57BL/6-8C5). Lineage FACS analysis data are plotted as the percentage of B220⁺, CD3⁺ and myeloid (Gr-1⁺, Mac-1⁺ and Gr-1⁺Mac-1⁺) cells among donor-derived Ly5.2⁺ cells in case of a transplantation experiment or among total white blood cells. As for early haematopoiesis analysis, mononuclear cells were isolated by low-density centrifugation (Histopaque 1083, Sigma) and stained with a cocktail of biotinylated lineage antibodies. Biotinylated antibodies used for lineage staining were all rat anti-mouse antibodies: anti-CD11b (clone M1/70), anti-B220 (clone RA3-6B2), anti-CD5 (clone 53-7.3), anti-Gr-1 (clone RB6-8C5), anti-Ter119 and anti-CD8a (clone 53-6.7) (all from eBioscience). After lineage depletion by magnetic separation (DynaBeads, Invitrogen), cells were stained with anti-Sca-1 (clone D7) (eBioscience), anti-c-Kit (clone 2B8) (eBioscience), anti-CD34 (clone RAM34) (eBioscience), anti-CD127 (clone A7R34) (eBioscience), anti-Flk-2 (clone A2F10) (eBioscience) and streptavidin (eBioscience). Early haematopoiesis FACS analysis data were plotted as percentage of long-term haematopoietic stem cells (LT-HSCs, gated as LSK CD34^{low}Flk2⁻), short-term haematopoietic stem cells (ST-HSCs, gated as LSK CD34⁺Flk2⁻) and lymphoid-primed multipotent progenitors (LMPPs, gated as LSK CD34⁺Flk2⁺)²⁸ distributed among donor-derived LSKs (Lin^{neg}c-Kit⁺Sca-1⁺ cells). To isolate LT-HSCs, lineage depletion was performed to enrich for lineage-negative cells. Lineage-negative cells were then stained as mentioned above and sorted using a BD FACS Aria III (BD Bioscience). For intracellular flow cytometric staining of β-catenin, lineage-depleted young, aged and young plus 100 ng ml⁻¹ Wnt5a bone marrow cells were incubated for 16 h in IMDM plus 10% FBS at 37 °C, 5% CO₂, 3% O₂. At the end of the treatment, the samples were moved on ice and stained again with the cocktail of biotinylated lineage antibodies. After washing, the samples were stained with anti-Sca-1 (clone D7) (eBioscience), anti-c-Kit (clone 2B8) (eBioscience), anti-CD34 (clone RAM34) (eBioscience), anti-Flk2 (clone A2F10) (eBioscience) and streptavidin (eBioscience). At the end of the surface staining, cells were fixed and permeabilized with Cytofix/Cytoperm solution (BD Biosciences) and incubated with 10% donkey serum (Sigma) in BD Perm/Wash Buffer (BD Biosciences) for 30 min. Primary and secondary antibody incubations were performed at room temperature in BD Perm/Wash Buffer (BD Biosciences) for 1 h and 30 min, respectively. The primary antibody for β-catenin was obtained from Millipore (rabbit polyclonal). The secondary antibody is a donkey anti-rabbit DyLight649 (BioLegend). Z-stacks were obtained by automatically scanning along the z axis of the cell with a confocal microscope and acquiring a picture of the in-focus plane every 0.6 μm.

Calcium flux protocol. Lin⁻Sca-1⁺c-Kit⁺ (LSK) cells were sorted out of 10 pooled 10-week-old male C57BL/6 mice. The sorted cells were suspended in 1 ml CLM in a

15-ml tube (CLM (cell loading medium): HBSS, 1% FCS, 1 mM CaCl₂, 1 mM MgCl₂). The cells were loaded with indo-1 AM (Molecular Probes I-1203, final concentration 0.25 μM) for 45 min at 37 °C in the dark. After loading, the cells were washed twice with CLM. The indo-1 cells were then resuspended in 1 ml CLM again and stored in the dark at room temperature for 1 h. Before flow cytometric analysis, the cells were equilibrated at 37 °C in the dark for 30 min. The cells were analysed by flow cytometry. An aliquot of the untreated cells was run to establish the baseline fluorescence of the indo-1-loaded cells. In one sample, ionomycin was used as positive control for Ca²⁺ release (1 μg ml⁻¹ final concentration). After a few minutes a calcium chelator, EGTA, was added during acquisition, and served as negative (Ca²⁺ low) control (8 mM final concentration). The response to Wnt5a was measured by adding murine recombinant Wnt5a (R&D Systems) to a final concentration of 300 ng ml⁻¹, 1 min after start of the measurement. Seven minutes after the first Wnt5a addition, a second, higher concentration of Wnt5a was added (700 ng ml⁻¹). To determine the calcium response to addition of Wnt5a, indo-1-loaded LSKs were analysed relative to a time parameter, and the change in fluorescence ratio over time can be related to changes in activation or stimulation by some agonist that will elicit a calcium influx. For visualization of this influx, the change in ratio of indo-1-bound Ca²⁺ (420 nm) and free indo-1 (510 nm) was depicted against the time after start of the measurement.

Immunofluorescence staining. Freshly sorted LT-HSCs were seeded on fibronectin-coated glass coverslips. For polarity staining, LT-HSCs were incubated for 12–16 h in HBSS + 10% FBS and when indicated treated with 100 ng ml⁻¹ Wnt5a (R&D System), casin (referred to in ref. 22 as Pirl1-related compound 2, obtained from Chembridge Corporation, and purified to greater than 99% by high-performance liquid chromatography) or left untreated. After incubation at 37 °C, 5% CO₂, 3% O₂, in growth factor-free medium, cells were fixed with BD Cytofix fixation buffer (BD Biosciences). After fixation cells were gently washed with PBS, permeabilized with 0.2% Triton X-100 (Sigma) in PBS for 20 min and blocked with 10% donkey serum (Sigma) for 30 min. Primary and secondary antibody incubations were performed for 1 h at room temperature. Coverslips were mounted with ProLong Gold Antifade reagent with or without DAPI (Invitrogen, Molecular Probes). The cells were stained with an anti-α-tubulin antibody (Abcam, rat monoclonal ab6160) detected with an anti-rat AMCA-conjugated secondary antibody or an anti-rat DyLight488-conjugated antibody (Jackson ImmunoResearch); an anti-Cdc42 antibody (Millipore, rabbit polyclonal) or an anti-β-catenin antibody (Millipore, rabbit polyclonal) detected with an anti-rabbit DyLight549-conjugated antibody (Jackson ImmunoResearch); an anti-pericentrin-2 antibody (Santa Cruz Biotechnology, goat polyclonal) detected with an anti-goat AMCA-conjugated antibody (Jackson ImmunoResearch); an anti-Wnt5a antibody (R&D System, goat polyclonal) detected with an anti-rat DyLight488-conjugated secondary antibody (Jackson ImmunoResearch). Samples were imaged with an AxioObserver Z1 microscope (Zeiss) equipped with a ×63 PH objective. Images were analysed with AxioVision 4.6 software. Alternatively, samples were analysed with an LSM710 confocal microscope (Zeiss) equipped with a ×63 objective. Primary raw data were imported into the Velocity Software package (Version 6.0, Perkin Elmer) for further processing and conversion into three-dimensional images. As for polarity scoring, the localization of each single stained protein was considered polarized when a clear asymmetric distribution was visible by drawing a line across the middle of the cell. A total of 50 to 100 LT-HSCs were singularly analysed per sample. Data are plotted as percentage of the total number of cells scored per sample. Specificity of the anti-Cdc42 antibody in immunofluorescence was tested on LT-HSCs sorted from mice in which *Cdc42* was deleted specifically in the haematopoietic system (Mx1-Cre;*Cdc42*^{lox/lox} mice²⁹) (data not shown).

Immunocytofluorescence microscopy. Freshly sorted LT-HSCs were seeded on fibronectin-coated glass coverslips. LT-HSCs were incubated for 2 h in HBSS + 10% FBS and when indicated treated with 100 ng ml⁻¹ Wnt5a (R&D System) or left untreated. After incubation at 37 °C, 5% CO₂, 3% O₂, in growth-factor-free medium cells were fixed with BD Cytofix Fixation buffer (BD Biosciences). Cells were incubated with primary antibodies diluted in blocking buffer overnight at 4 °C. Cells were then washed three times with PBS and incubated with a secondary antibody diluted 1:1,000 in blocking buffer overnight at 4 °C. The primary and secondary antibodies used were: anti-CamkII (polyclonal rabbit, Cell Signaling Technology), phosphor-Thr 286-CamkII (polyclonal rabbit, Cell Signaling Technology), and anti-NFATc1 (mouse monoclonal IgG1, clone 7A6, Santa Cruz Biotechnology). After final washes, the coverslips were mounted in SlowFade Gold Antifade reagent supplemented with DAPI (4,6-diamino-2-phenylindole, dihydrochloride) nuclear stain (Invitrogen). Fluorescence digital images were taken using constant settings on a Leica DM RBE fluorescent microscope (Leica) using AxioVision software (Carl Zeiss). For each particular sample, images at ×100 magnification of between 30 and 100 randomly captured cells were taken. Fluorescence digital images were then analysed using the digital image processing software ImageJ (NIH). The mean fluorescence intensity (average intensity of pixels per cell) was determined.

Background correction was performed by determining the signal intensity of the pixels around the perimeter of the area being quantified. These background signals were then subtracted from specific signals caused by antibody staining. To compare measurements from separate experiments, they were additionally normalized to the mean of a set of control samples and expressed as fold changes in relation to the control samples. For comparison of means of different groups, a two-tailed *t*-test assuming equal variances was used. A *P* value less than 0.05 was considered to be statistically significant.

Western blot and Cdc42-GTPase effector domain pull-down assays. Relative levels of GTP-bound Cdc42 were determined by an effector pull-down assay. Briefly, lineage-depleted bone-marrow cells (10^6) were lysed in a Mg^{2+} lysis/wash buffer (Upstate cell signalling solutions) containing 10% glycerol, 25 mM sodium fluoride, 1 mM sodium orthovanadate and a protease inhibitor cocktail (Roche Diagnostics). Samples were incubated with PAK-1 binding domain/agarose beads and bound (activated) as well as unbound (non-activated) Cdc42 fractions were probed by immunoblotting with an anti-Cdc42 antibody (Millipore, rabbit polyclonal). Activated protein was normalized to total protein and/or β -actin (Sigma) and the relative amount was quantified by densitometry. For detection of Wnt5a protein levels, cells were lysed as mentioned above and directly blotted with a goat anti-mouse Wnt5a antibody (R&D Systems). Wnt5a protein levels were normalized based on actin protein levels in the same blotted samples.

Reverse-transcriptase real-time PCR. 20,000–40,000 LT-HSCs from young and aged mice were lysed and processed for RNA extraction immediately after sorting. RNA was obtained with the microRNA Extraction kit (Qiagen) and all was used for cDNA conversion. cDNA was prepared and amplified with Ovation RNA Amplification system V2 (NuGEN). All real-time PCRs were run with TaqMan real-time PCR reagent and primers from Applied Biosystem on an ABI 9700HT real time machine.

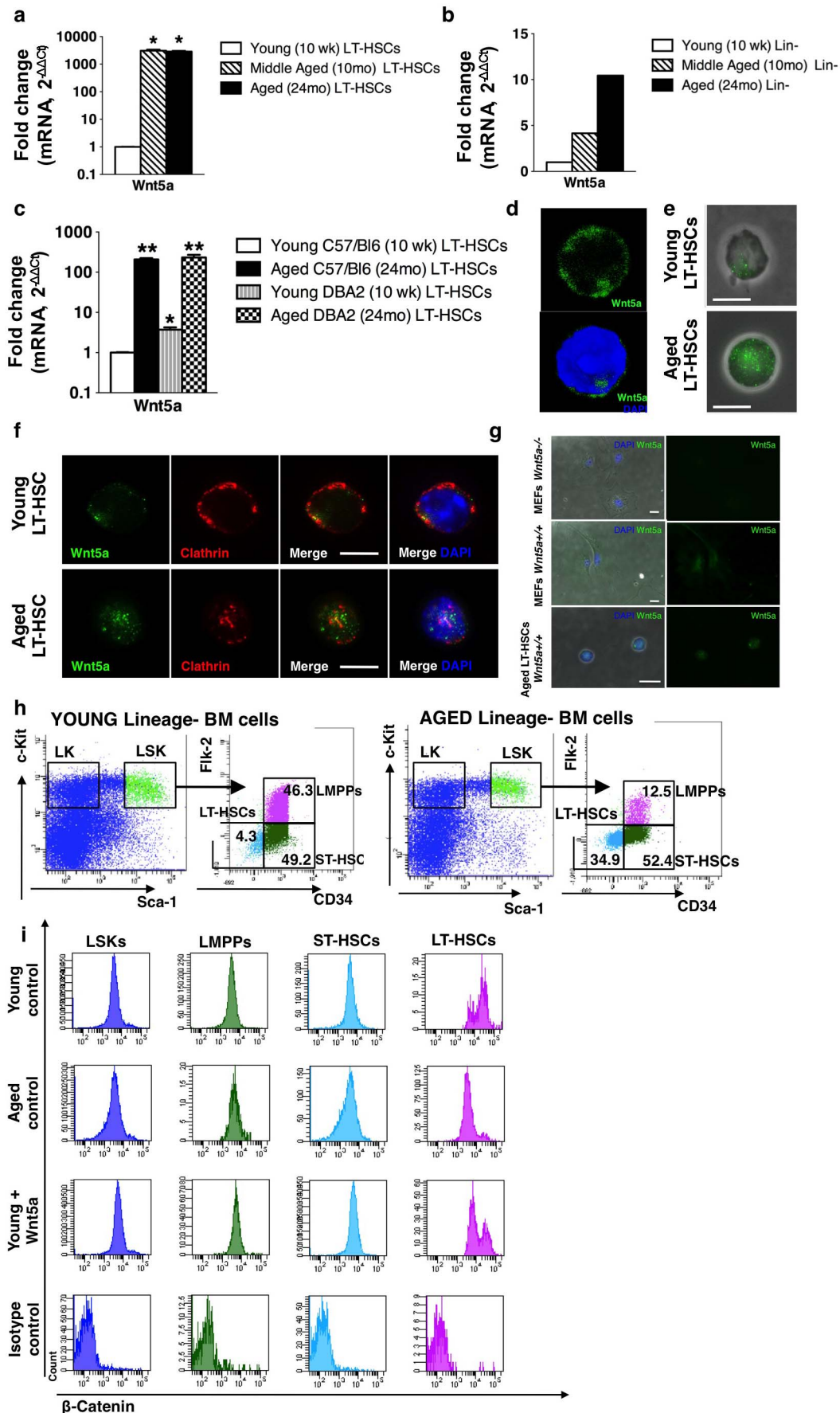
Stroma CD45⁺ cells. To isolate cells close to the endosteum, femora and tibiae were isolated from young (2–3-month-old) and aged (22–23-month-old) mice. The bones were cleaned and the associated muscle tissues removed. After the bone marrow was flushed, the bones were crushed using scissors and minced with a scalpel in 1.5 mg ml^{-1} collagenase IV (Worthington)/PBS. The bone chips were further incubated and shaken for 1.5 h at 37°C in collagenase IV/PBS. The bone chips were washed extensively with IMDM/10%FBS and the dissociated cells collected. This stroma cell fraction was filtered through a $100\text{ }\mu\text{m}$ cell strainer and stained for with an anti-CD45 antibody (clone 104), and CD45 negative cells sorted by flow cytometry.

shRNA lentiviral transduction. Aged mice (24-month-old) were killed; bone marrow mononuclear cells were isolated by low-density centrifugation (Histopaque 1083, Sigma) and stained with a cocktail of biotinylated lineage antibodies. Biotinylated antibodies used for lineage staining were all rat anti-mouse antibodies: anti-CD11b (clone M1/70), anti-B220 (clone RA3-6B2), anti-CD5 (clone 53-7.3) anti-Gr-1 (clone RB6-8C5), anti-Ter119 and anti-CD8a (clone 53-6.7) (all from eBioscience). Not pre-stimulated, aged, lineage-depleted bone marrow cells (Lin[−]BM) were transduced overnight on retronectin-coated (TaKaRa) plates with cell-free

supernatants containing lentiviral particles according to refs 30, 31. The lentivirus plasmid vector pLKO.1-YFP was obtained from Sigma's validated genome-wide TRC shRNA libraries (Sigma-Aldrich).

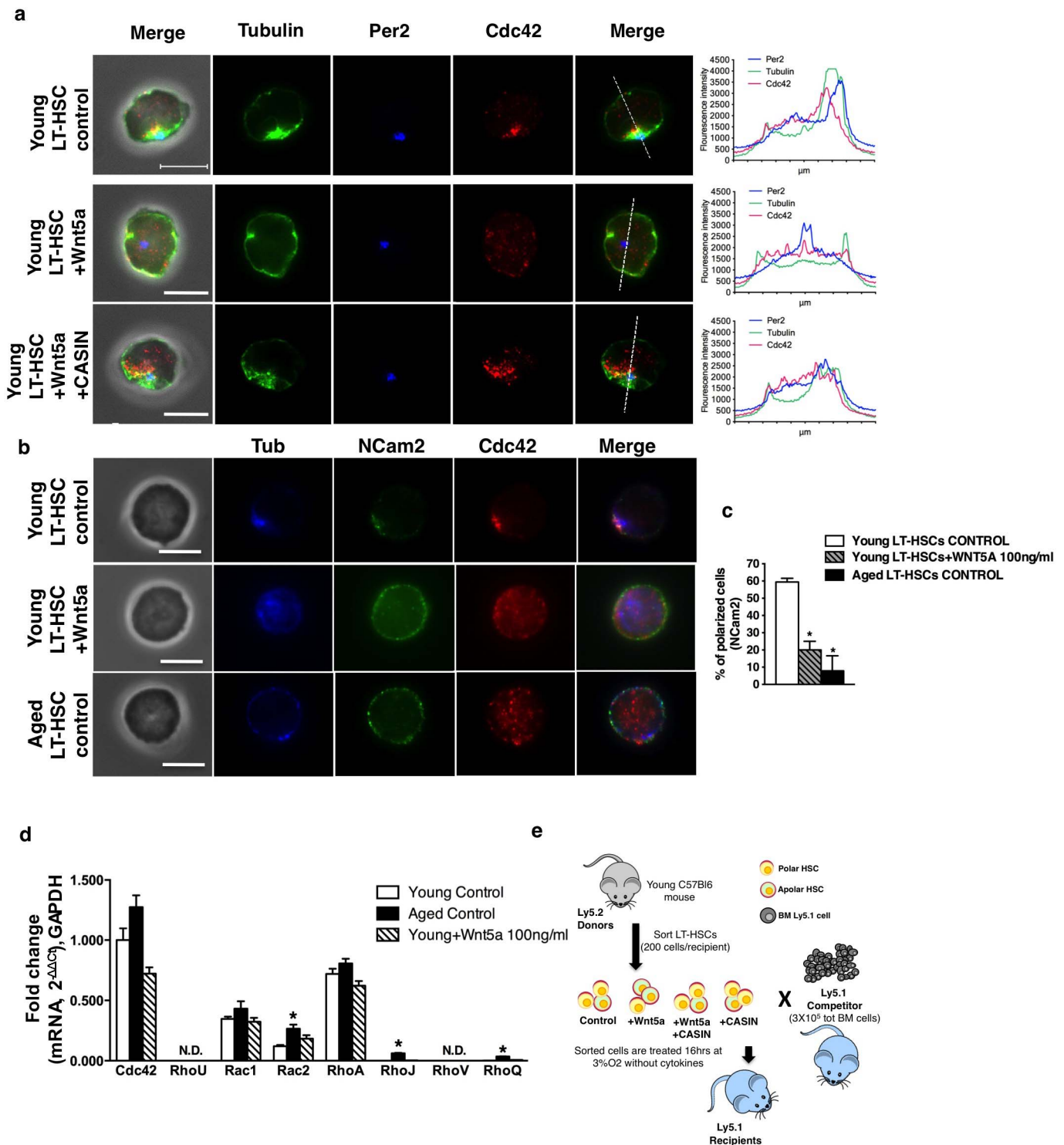
Statistical analyses. Data were assumed to meet normal distribution. The variance was similar between groups that were statistically compared. All data are plotted as mean + 1 standard error (s.e.m.) unless differently stated. The s.e.m. is used to indicate the precision of an estimated mean. Such a data representation does not affect the statistical analyses as variance information is used in the test statistics. A paired Student's *t*-test was used to determine the significance of the difference between means of two groups. One-way ANOVA or two-way ANOVA were used to compare means among three or more independent groups. Bonferroni post-test to compare all pairs of data set was determined when overall *P* value was <0.05 . All statistical analyses were determined with Prism 4.0c version. To choose sample size, we used GraphPad StatMate Software Version 2.0b, estimating a standard deviation between 2 and 8 (depending on the experiment and the possibility of increasing sample size). For transplantation experiments we estimated a sample size of 15–20 (assuming a standard deviation of 10 and a significant difference between means of at least 15). In transplantation experiments, samples were included in the analysis when engraftment was more or equal to 1.0% after at least 12 weeks from injections (24 weeks for most of the experiments) and contribution was detected in all lineages. Mice showing signs of sickness and with clear alterations of blood parameter and/or showing signs of major disease involving also non-haematopoietic tissues were excluded from analysis. As for *in vitro* experiments, samples were excluded from analysis in case of clear technical problems (error in immune-blotting or staining procedures or technical problems with reagents). All criteria for exclusions of samples from *in vivo* or *in vitro* experiments were pre-established. In each figure legend, the number (*n*) of biological repeats (samples obtained from experiments repeated in different days and starting from different mice) included in the final statistical analysis is indicated. Mice for experiments were randomly chosen from our in-house colonies or suppliers. All mice were C57BL/6 females unless differently stated. The investigator was not blinded to the mouse group allocation nor when assessing the outcome (aged mice or young mice transplanted with aged bone marrow stem cells require particular care and follow up).

28. Adolfsson, J. *et al.* Identification of Flt3⁺ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
29. Yang, L. *et al.* Cdc42 critically regulates the balance between myelopoiesis and erythropoiesis. *Blood* **110**, 3853–3861 (2007).
30. Daria, D. *et al.* The retinoblastoma tumor suppressor is a critical intrinsic regulator for hematopoietic stem and progenitor cells under stress. *Blood* **111**, 1894–1902 (2008).
31. Li, Z. *et al.* Predictable and efficient retroviral gene transfer into murine bone marrow repopulating cells using a defined vector dose. *Exp. Hematol.* **31**, 1206–1214 (2003).



Extended Data Figure 1 | Increased expression of *Wnt5a* in aged LT-HSCs results in a shift from canonical to non-canonical Wnt signalling. **a**, Reverse transcriptase real-time PCR analysis of *Wnt5a* transcript levels in young (10-week-old), middle-aged (10-month-old) and aged (24-month-old) LT-HSCs ($\text{Lin}^- \text{c-kit}^+ \text{Sca-1}^+ \text{Flk2}^- \text{CD34}^-$ bone marrow cells) sorted from C57BL/6 mice. Data are expressed as fold increased compared to the lowest expressed transcript arbitrarily set to 1. *Wnt5a* mRNA is barely detectable in young LT-HSCs and is markedly upregulated in middle-aged and aged LT-HSCs. Data were analysed with the $2^{-\Delta\Delta C_t}$ method and plotted on a logarithmic scale. Bars are mean \pm 1 s.e.; $n = 3$, $*P < 0.05$. **b**, Reverse transcriptase real-time PCR analysis of *Wnt5a* transcript levels in young (10-week-old), middle-aged (10-month-old) and aged (24-month-old) Lin^- bone marrow cells from C57BL/6 mice. Data are expressed as fold increased compared to the lowest expressed transcript arbitrarily set to 1. *Wnt5a* mRNA is barely detectable in young Lin^- cells and is upregulated in middle-aged and aged LT-HSCs. Data were analysed with the $2^{-\Delta\Delta C_t}$ method and plotted on a logarithmic scale. Scale bars represent results of one set of samples. The experiment was repeated twice with similar results. **c**, Reverse transcriptase real-time PCR analysis of *Wnt5a* transcript levels in young (10-week-old) and aged (24-month-old) LT-HSCs ($\text{Lin}^- \text{c-Kit}^+ \text{Sca-1}^+ \text{Flk2}^- \text{CD34}^-$ bone marrow cells) sorted from C57BL/6 and DBA/2 mice. Data are expressed as fold increased compared to the lowest expressed transcript arbitrarily set to 1. Data were analysed with the $2^{-\Delta\Delta C_t}$ method and plotted on a logarithmic scale. Error bars are mean \pm 1 s.e.; $n = 3$, $*P < 0.05$, $**P < 0.01$. **d**, Representative

three-dimensional confocal picture of *Wnt5a* distribution in an aged LT-HSC. The nucleus is stained with DAPI. Three-dimensional localization of *Wnt5a* was analysed by scanning the cells along the z-axis and acquiring a picture of the xy-plane every 0.7 μm . Three-dimensional images were then reconstructed by using Volocity v6.0 software. **e**, Representative immunofluorescence picture of *Wnt5a* (green) membrane distribution (non-permeabilized cells) in young and aged LT-HSCs. Immunofluorescence pictures are shown as overlap with the phase contrast image. Scale bar, 5 μm . **f**, Representative immunofluorescence picture of *Wnt5a* (green) and clathrin (red) localization in young and aged LT-HSCs. Pictures are shown on a dark background and as overlap with DAPI (staining nuclei). Scale bar, 5 μm . **g**, Representative expression of *Wnt5a* in MEFs (mouse embryonic fibroblasts) and aged LT-HSCs from *Wnt5a*^{+/+} mice determined by immunofluorescence. *Wnt5a* fluorescence signal is not detected when MEFs from *Wnt5a*^{-/-} mice are stained with the same procedure. *Wnt5a* pictures are shown on a dark background and as overlap with DAPI (blue, staining nuclei) and phase contrast images. Scale bar, 10 μm . **h**, Representative FACS dot plots of LT-HSCs ($\text{Lin}^- \text{c-Kit}^+ \text{Sca-1}^+ \text{Flk2}^- \text{CD34}^-$), ST-HSCs ($\text{Lin}^- \text{c-Kit}^+ \text{Sca-1}^+ \text{Flk2}^- \text{CD34}^+$), LMPPs ($\text{Lin}^- \text{c-Kit}^+ \text{Sca-1}^+ \text{Flk2}^+ \text{CD34}^+$), LSKs ($\text{Lin}^- \text{c-Kit}^+ \text{Sca-1}^+$) and LKs ($\text{Lin}^- \text{c-Kit}^+ \text{Sca-1}^-$) gating strategy of young and aged lineage-depleted bone marrow cells. **i**, Representative FACS histograms of β -catenin expression in young, aged and young *Wnt5a*-treated LT-HSCs, ST-HSCs, LMPPs and LSKs.

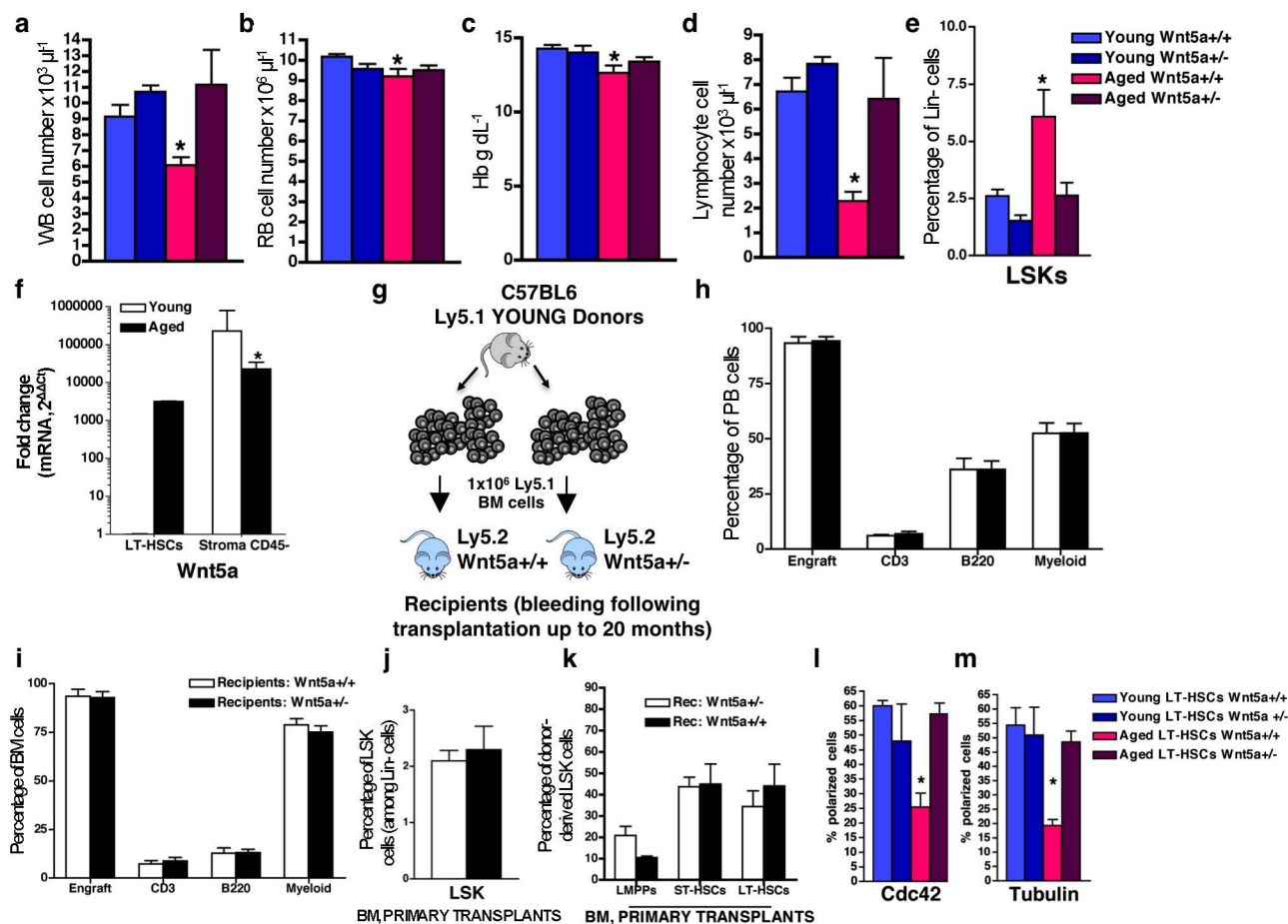


Extended Data Figure 2 | Wnt5a activates Cdc42 inducing ageing-like phenotypes in young LT-HSCs. **a**, Representative distribution of Cdc42, tubulin and Per2 (staining the centrosome) in young control, young Wnt5a-treated (100 ng ml^{-1}) and young Wnt5a (100 ng ml^{-1}) + casin ($5 \mu\text{M}$)-treated LT-HSCs determined by immunofluorescence. Scale bar, $5 \mu\text{m}$. Shown are also representative fluorescence intensity plots obtained by collecting pixel intensity through the section of the cell as indicated by the dotted line in the corresponding merge picture. **b**, Representative distribution of Cdc42, tubulin and NCAM2 (membrane protein) in young control, young Wnt5a-treated (100 ng ml^{-1}) and aged LT-HSCs determined by immunofluorescence. Scale bar, $5 \mu\text{m}$. **c**, Graph of the percentage of young control, young Wnt5a-treated

(100 ng ml^{-1}) and aged LT-HSCs with a polar distribution of NCAM2. Shown are mean \pm 1 s.e., $n = 4$; ~ 200 – 300 LT-HSCs scored per sample in total.

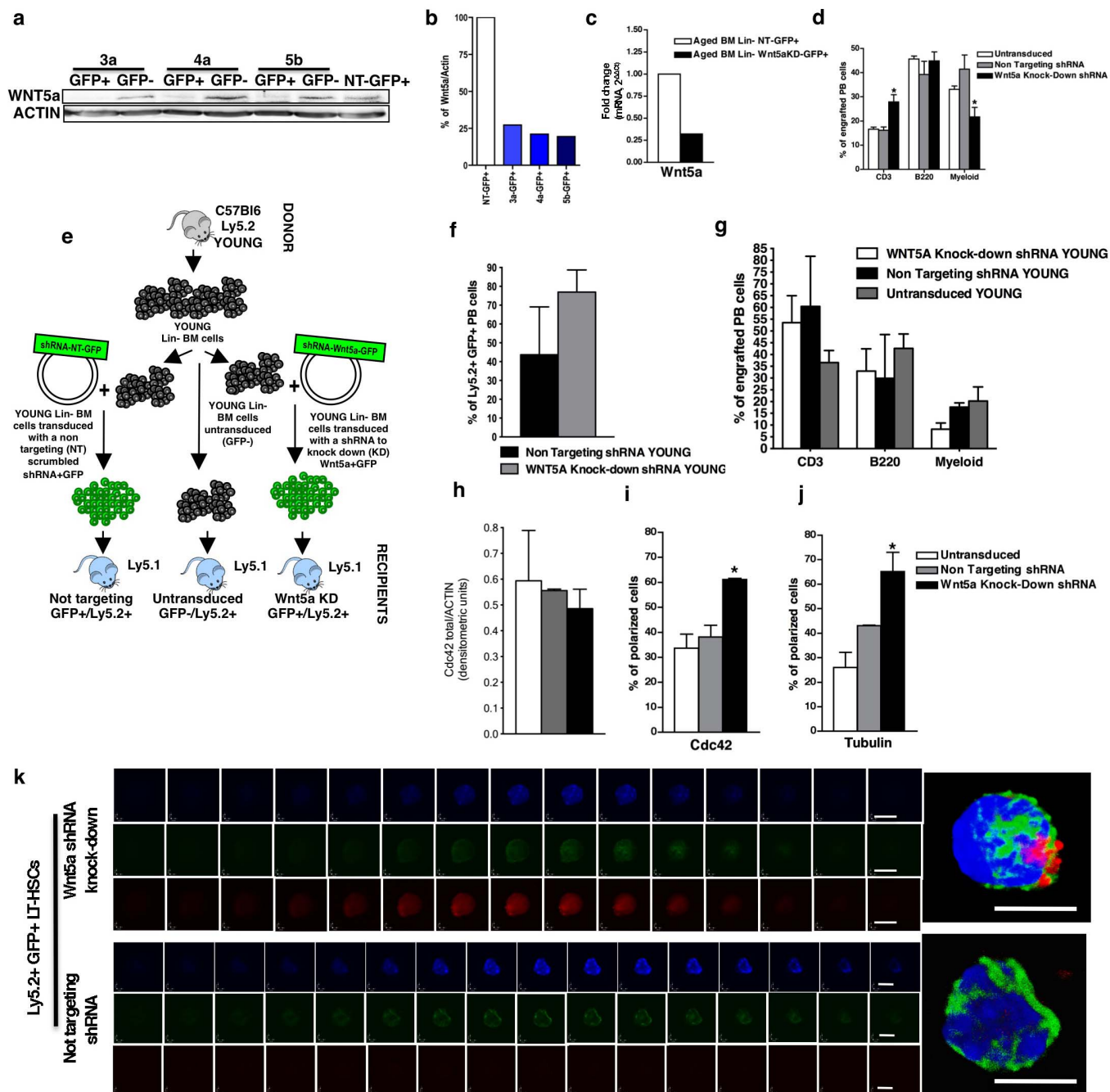
$*P < 0.001$. **d**, Reverse transcriptase real-time PCR analysis of *Cdc42*, *Rhou*, *Rac1*, *Rac2*, *Rhoa*, *Rhoj*, *Rhov* and *Rhoq* transcript levels in young, aged and young Wnt5a-treated (100 ng ml^{-1} , 16 h treatment) LT-HSCs. *Rhou* and *Rhov* transcripts were below detection limits in all the assayed samples (ND). Data are expressed as fold difference compared to the expression of *Cdc42* mRNA in young LT-HSCs arbitrarily set to 1. Data were analysed with the $2^{-\Delta\Delta C_t}$ method and plotted on a linear scale. Bars are mean \pm 1 s.e.; $n = 3$, $*P < 0.05$.

e, Schematic representation of the experimental set-up for the transplantation. Recipient mice were analysed 24 weeks after transplant.



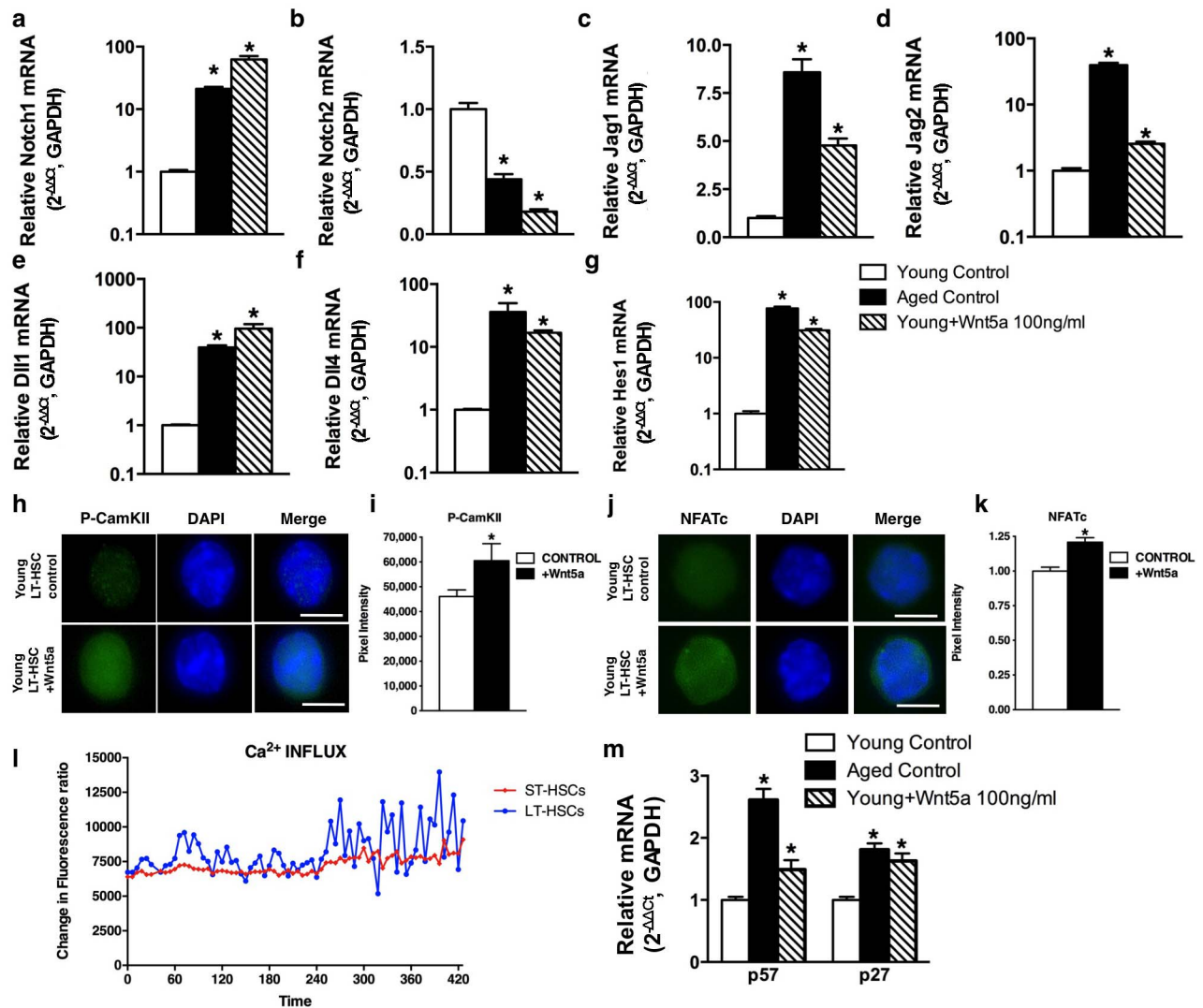
Extended Data Figure 3 | *Wnt5a* haploinsufficient mice present with attenuated HSC ageing. **a–d**, White blood (WB) cell count (**a**), red blood (RB) cell count (**b**), haemoglobin (Hb) dosage (**c**) and lymphocyte cell count (**d**) in peripheral blood of young and aged *Wnt5a*^{+/-} and *Wnt5a*^{+/+} mice. **P* < 0.05; shown are mean + 1 s.e., *n* = 5. **e**, Percentage of LSKs among Lin⁺ cells in bone marrow of young and aged *Wnt5a*^{+/-} and *Wnt5a*^{+/+} mice. **P* < 0.05; shown are mean + 1 s.e., *n* = 5. **f**, Reverse transcriptase real-time PCR analysis of *Wnt5a* transcript levels in young and aged LT-HSCs and young and aged collagenase-digested and sorted CD45⁺ cells (stroma cells). Data are expressed as fold difference compared to the expression in young LT-HSCs arbitrarily set to 1. *Wnt5a* mRNA shows significantly increased expression in stroma CD45⁺ cells when compared to young and aged LT-HSCs. In contrast to the situation in LT-HSCs, young stroma CD45⁺ cells express higher levels of *Wnt5a* mRNA than aged stroma CD45⁺ cells. Data were analysed with the $2^{-\Delta\Delta\text{CT}}$ method and plotted on a logarithmic scale. Error bars are mean + 1 s.e.; *n* = 4, **P* < 0.05. **g**, Schematic representation of the experimental set-up for

transplantation. Young donor (Ly5.1⁺) bone marrow cells were transplanted into recipient (Ly5.2⁺) young *Wnt5a*^{+/-} and *Wnt5a*^{+/+} mice. Recipient mice were killed and analysed 20 months after transplant. **h, i**, Percentage of engrafted cells, B220⁺, CD3⁺ and myeloid cells among donor-derived Ly5.1⁺ cells in peripheral blood (**h**) and bone marrow (**i**) 20 months after transplants. Columns are mean values + 1 s.e., *n* = 5. **j**, Percentage of donor-derived LSKs among donor-derived Lin⁺ cells in bone marrow of *Wnt5a*^{+/-} and *Wnt5a*^{+/+} recipient mice 20 months after transplant. Columns show mean + 1 s.e.m., *n* = 5. **k**, Percentage of donor-derived LT-HSCs, ST-HSCs and LMPs among donor-derived LSKs in *Wnt5a*^{+/-} and *Wnt5a*^{+/+} recipient mice 20 months after transplant. Columns are mean values + 1 s.e. *n* = 5. **l, m**, Percentage of LT-HSCs polarized for Cdc42 (**l**) and tubulin (**m**) in young and aged *Wnt5a*^{+/-} and *Wnt5a*^{+/+} mice. Shown are mean + 1 s.e., *n* = 4 and 200 cells scored per sample in total. **P* < 0.01 versus young *Wnt5a*^{+/+} and *P* < 0.05 versus young *Wnt5a*^{+/-} and aged *Wnt5a*^{+/-} mice.



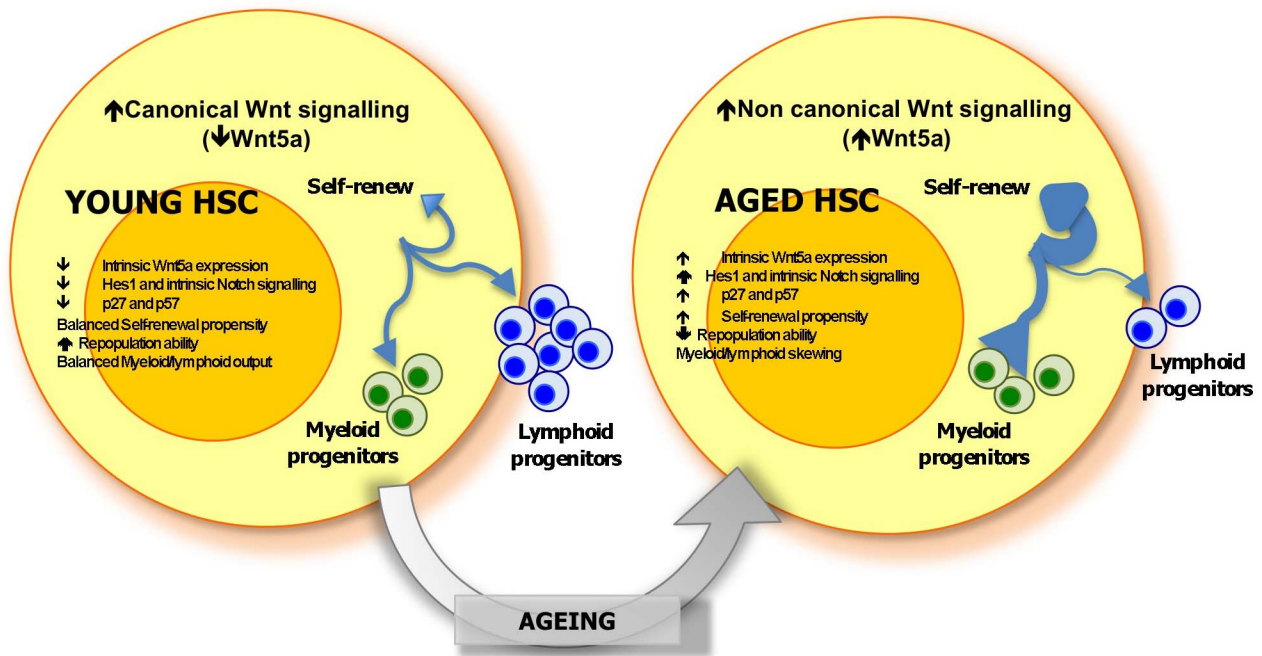
Extended Data Figure 4 | Validation of the knockdown efficiency in 3T3 fibroblast cells and aged and young Lin^- bone marrow cells. **a**, Transduced fibroblast cells were sorted and analysed by western blot for Wnt5a protein levels. Wnt5a protein levels were normalized on actin. Three different Wnt5a knockdown vectors (3a-GFP⁺, 4a-GFP⁺, 5b-GFP⁺) were tested and Wnt5a protein levels were compared to non-targeting transduced fibroblasts (NT-GFP⁺) and to untransduced cells sorted as GFP⁻ from the initial mixed culture (3a-GFP⁻, 4a-GFP⁻, 5b-GFP⁻). **b**, Transduced fibroblast cells were sorted and analysed by reverse transcriptase real-time PCR for *Wnt5a* mRNA levels. *Wnt5a* mRNA levels are normalized to actin mRNA levels. Three different Wnt5a knockdown vectors (3a-GFP⁺, 4a-GFP⁺, 5b-GFP⁺) were tested and *Wnt5a* transcript levels were compared to non-targeting transduced fibroblasts (NT-GFP⁺). **c**, Not pre-stimulated transduced Wnt5a^{KD} or non-targeted aged Lin^- bone marrow cells were sorted and analysed by reverse transcriptase real-time PCR for *Wnt5a* mRNA levels. *Wnt5a* mRNA levels are normalized on *Gapdh* mRNA levels. **d**, Percentage of B220⁺, CD3⁺ and myeloid cells among donor-derived cells in peripheral blood 24 weeks after transplant. * $P < 0.05$; shown are mean values + 1 s.e. Mice were considered as engrafted when the percentage of Ly5.2⁺ cells in peripheral blood was higher than 1.0 and contribution was detected for all peripheral blood lineages. Data are based on two different lentiviral infection/transplant experiments with 5–7 recipient mice per group (for example, $n = 10$). **e**, Schematic representation of the experimental set-up for transplantation of Wnt5a knock down (Wnt5a^{KD}), Wnt5a non-targeting (Wnt5a-NT) and untransduced young haematopoietic progenitor/stem cells. Young donor (Ly5.2⁺) lineage-negative (Lin^-) bone marrow cells were infected with the indicated lentiviral vectors or

left untransduced. Infected cells were sorted based on GFP expression. Cells ($1\text{--}3 \times 10^5$ Ly5.2⁺) were transplanted into recipients (Ly5.1⁺). Recipient mice were analysed 12–16 weeks after transplant. **f**, Percentage of engrafted donor-derived cells in peripheral blood 12–16 weeks after transplant. Shown are mean values + 1 s.e. Mice were considered as engrafted when the percentage of Ly5.2⁺ cells in peripheral blood was higher than 1.0 and contribution was detected for all peripheral blood lineages. Data are based on two different lentiviral infection/transplant experiments with 3 recipient mice per group (for example, $n = 3$ for Wnt5a^{KD} and Wnt5a-NT mice and $n = 6$ for untransduced mice). **g**, Percentage of B220⁺, CD3⁺ and myeloid cells among donor-derived cells in peripheral blood 24 weeks after transplant. * $P < 0.05$; shown are mean values + 1 s.e. Mice were considered as engrafted when the percentage of Ly5.2⁺ cells in peripheral blood was higher than 1.0 and contribution was detected for all peripheral blood lineages. Data are based on two different lentiviral infection/transplant experiments with 5–7 recipient mice per group (for example, $n = 10$). **h**, Ratio of the densitometric score of the total Cdc42 expression as shown in Fig. 4d. The experiment was repeated four times with mice (1 mouse for 1 sample) from different lentiviral infection/transplant experiments. Shown are mean + 1 s.e., $n = 4$, * $P < 0.05$. **i**, **j**, Percentage of donor-derived LT-HSCs polarized for Cdc42 (**i**) and tubulin (**j**) 24 weeks after transplant. Shown are mean values + 1 s.e., $n = 4$, ~200 cells scored per sample in total. * $P < 0.05$. **k**, Representative immunofluorescence z-stack pictures of tubulin (green) and β -catenin (red) localization in aged Wnt5a-NT (Ly5.2⁺GFP⁺) or aged Wnt5a^{KD} (Ly5.2⁺GFP⁺) LT-HSCs. Nuclei are stained with DAPI (blue). Shown is also the final three-dimensional reconstructed merged image. Scale bar, 5 μm .



Extended Data Figure 5 | Wnt pathways in HSCs and ageing. **a–g**, Reverse transcriptase real-time PCR analysis of *Notch1*, *Notch2*, *Jag1*, *Jag2*, *Dll1* (delta like1), *Dll4* (delta like 4) and *Hes1* transcript levels in young, aged and young Wnt5a-treated (16 h treatment) LT-HSCs. *Notch3*, *Notch4* and *Dll3* (delta like 3) transcripts were below detection limits in all the assayed samples. Data are expressed as fold difference compared to the expression in young LT-HSCs arbitrarily set to 1. Data were analysed with the $2^{-\Delta\Delta C_t}$ method and plotted on a logarithmic or linear scale. Bars are mean \pm 1 s.e.; $n = 3$, $*P < 0.05$. **h**, Representative immunofluorescence picture of p-CamKII (green) expression and localization in young control and young Wnt5a-treated LT-HSCs. Pictures are shown on a dark background and as overlap with DAPI (staining nuclei). Scale bar, 5 μ m. **i**, Relative expression of p-CamKII in young LT-HSCs and on Wnt5a treatment, determined by integration of pixel intensity. $*P < 0.05$.

j, Representative immunofluorescence picture of NFATc (green) expression and localization in young control and young Wnt5a-treated LT-HSCs. Pictures are shown on a dark background and as overlap with DAPI (staining nuclei). Scale bar, 5 μ m. **k**, Relative expression of NFATc in young LT-HSCs and on Wnt5a treatment, determined by integration of pixel intensity. $*P < 0.05$. **l**, Changes in intracellular Ca^{2+} concentrations in ST-HSCs and LT-HSCs in response to stimulation with Wnt5a as determined by flow cytometry. **m**, Reverse transcriptase real-time PCR analysis of p57 and p27 transcript levels in young, aged and young Wnt5a-treated (100 ng ml $^{-1}$, 16 h treatment) LT-HSCs. Data are expressed as fold difference compared to the expression in young LT-HSCs arbitrarily set to 1. Data were analysed with the $2^{-\Delta\Delta C_t}$ method and plotted on a logarithmic or linear scale. Bars are mean \pm 1 s.e.; $n = 3$, $*P < 0.05$.



Extended Data Figure 6 | Mechanisms of haematopoietic stem-cell ageing. Cartoon scheme summarizing the main phenotypic and functional differences between young and aged LT-HSCs.

Staphylococcus δ -toxin induces allergic skin disease by activating mast cells

Yuumi Nakamura¹, Jon Oscherwitz^{2,3}, Kemp B. Cease^{2,3}, Susana M. Chan¹, Raul Muñoz-Planillo¹, Mizuho Hasegawa¹, Amer E. Villaruz⁴, Gordon Y. C. Cheung⁴, Martin J. McGavin⁵, Jeffrey B. Travers⁶, Michael Otto⁴, Naohiro Inohara¹ & Gabriel Núñez¹

Atopic dermatitis is a chronic inflammatory skin disease that affects 15–30% of children and approximately 5% of adults in industrialized countries¹. Although the pathogenesis of atopic dermatitis is not fully understood, the disease is mediated by an abnormal immunoglobulin-E immune response in the setting of skin barrier dysfunction². Mast cells contribute to immunoglobulin-E-mediated allergic disorders including atopic dermatitis³. Upon activation, mast cells release their membrane-bound cytosolic granules leading to the release of several molecules that are important in the pathogenesis of atopic dermatitis and host defence⁴. More than 90% of patients with atopic dermatitis are colonized with *Staphylococcus aureus* in the lesional skin whereas most healthy individuals do not harbour the pathogen⁵. Several staphylococcal exotoxins can act as superantigens and/or antigens in models of atopic dermatitis⁶. However, the role of these staphylococcal exotoxins in disease pathogenesis remains unclear. Here we report that culture supernatants of *S. aureus* contain potent mast-cell degranulation activity. Biochemical analysis identified δ -toxin as the mast cell degranulation-inducing factor produced by *S. aureus*. Mast cell degranulation induced by δ -toxin depended on phosphoinositide 3-kinase and calcium (Ca^{2+}) influx; however, unlike that mediated by immunoglobulin-E crosslinking, it did not require the spleen tyrosine kinase. In addition, immunoglobulin-E enhanced δ -toxin-induced mast cell degranulation in the absence of antigen. Furthermore, *S. aureus* isolates recovered from patients with atopic dermatitis produced large amounts of δ -toxin. Skin colonization with *S. aureus*, but not a mutant deficient in δ -toxin, promoted immunoglobulin-E and interleukin-4 production, as well as inflammatory skin disease. Furthermore, enhancement of immunoglobulin-E production and dermatitis by δ -toxin was abrogated in *Kit*^{W^{sh}/W^{sh}} mast-cell-deficient mice and restored by mast cell reconstitution. These studies identify δ -toxin as a potent inducer of mast cell degranulation and suggest a mechanistic link between *S. aureus* colonization and allergic skin disease.

Because mast cells (MCs) may play a critical role in the pathogenesis of atopic dermatitis³, we asked first whether *S. aureus* can release factors that induce MC degranulation. We found that the culture supernatant of *S. aureus* induced rapid and robust MC degranulation in a dose-dependent manner (Fig. 1a and Supplementary Fig. 1a, b). Analysis of a panel of *Staphylococcus* isolates showed that the culture supernatant of several *S. aureus* strains as well as of that from *Staphylococcus epidermidis* and *Staphylococcus saprophyticus*, but not of several *Staphylococcus* species, elicited MC degranulation (Supplementary Fig. 1c). Toll-like receptor 2 (TLR2) stimulation by lipopeptides has been shown by some studies, but not others, to induce MC degranulation^{7,8}. However, neither the culture supernatant of *S. aureus* deficient in lipoproteins (*AlgT*), which lacks TLR2-stimulating activity⁹, nor that from bacteria deficient in α -, β - and γ -haemolysins ($\Delta\alpha\beta\gamma$) were impaired in MC degranulation activity

(Supplementary Figs 1c and 3c). The MC degranulation activity was enriched in the culture supernatant of *S. aureus* and was sensitive to heat, phenol/chloroform extraction and protease K treatment (Supplementary Fig. 2a). Furthermore, the MC degranulation-inducing factor bound to both diethylaminoethyl and carboxymethyl cellulose matrices and was present in the void fraction on gel filtration at neutral pH (Supplementary Fig. 2b). On the basis of these observations, we developed a many-step strategy for biochemical purification of the MC degranulation-inducing factor (Supplementary Fig. 2c). Liquid chromatography–mass spectrometry analysis showed that δ -toxin (also called δ -haemolysin or phenol-soluble modulins (PSM)- γ), a 2.9 kDa peptide secreted by *S. aureus* that belongs to the peptide toxin family of PSMs, was the most abundant and significant protein identified in the purified sample (Supplementary Fig. 2c). Mutant analyses in two *S. aureus* strains showed that MC degranulation induced by *S. aureus* culture supernatant required expression of δ -toxin whereas deficiency of related PSM- α or PSM- β peptides had minimal or no effect on MC degranulation (Fig. 1b and Supplementary Fig. 3a). Complementation of the *Δhld* mutant strain with δ -toxin-producing plasmid, but not control plasmid, restored the ability of the culture supernatant to induce MC degranulation (Fig. 1b). Stimulation of MCs with 30 $\mu\text{g ml}^{-1}$ of synthetic δ -toxin peptide, a concentration of δ -toxin normally found in *S. aureus* culture supernatants (Supplementary Fig. 3b), also induced rapid release of histamine (Fig. 1c). Furthermore, transmission electron microscopy showed classic features of MC degranulation without loss of plasma membrane integrity upon δ -toxin stimulation (Fig. 1d). These results indicate that δ -toxin is the MC degranulation-inducing factor released by *S. aureus*.

PSMs, especially PSM- α 2 and PSM- α 3, induce cell death and interleukin (IL)-8 release in human neutrophils^{10,11}. In accord with these results¹⁰, PSM- α 2 and PSM- α 3 induced robust loss of cell viability in MCs (Supplementary Fig. 4a). Non-toxic concentrations of PSM- α s did not possess any MC-degranulation activity (Supplementary Fig. 4b). In contrast, stimulation with a concentration of δ -toxin that induces robust MC degranulation did not induce detectable cell death in MCs (Supplementary Fig. 4a, c). Furthermore, formylation of the amino (N) terminus of the δ -toxin peptide was not required for MC degranulation activity, whereas it was essential for the ability of δ -toxin to induce the release of IL-8 from human neutrophils (Supplementary Fig. 4c, d). Consistent with previous results, stimulation of human neutrophils with formylated PSM- α 2, PSM- α 3 or δ -toxin induced robust IL-8 release (Supplementary Fig. 4d). Moreover, stimulation of primary mouse macrophages and keratinocytes with PSM- α 2, but not δ -toxin, triggered robust cell death (Supplementary Fig. 5). Thus, the MC degranulation activity induced by δ -toxin is not associated with cell death and is different from other activities triggered by PSM- α 2 and PSM- α 3. Immunoblotting confirmed that the presence of δ -toxin in *S. aureus* supernatants correlated with MC degranulation activity (Fig. 1e). Notably, the supernatant

¹Department of Pathology and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ²Department of Internal Medicine, Division of Hematology/Oncology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ³VA Ann Arbor Healthcare System, 2215 Fuller Road, Ann Arbor, Michigan 48105, USA. ⁴Laboratory of Human Bacterial Pathogenesis, National Institute of Allergy and Infectious Diseases, US National Institutes of Health, Bethesda, Maryland 20892, USA. ⁵Department of Microbiology and Immunology, and Centre for Human Immunology, Western University, University of Western Ontario, London, Ontario, N6A 5C1, Canada. ⁶Department of Dermatology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.

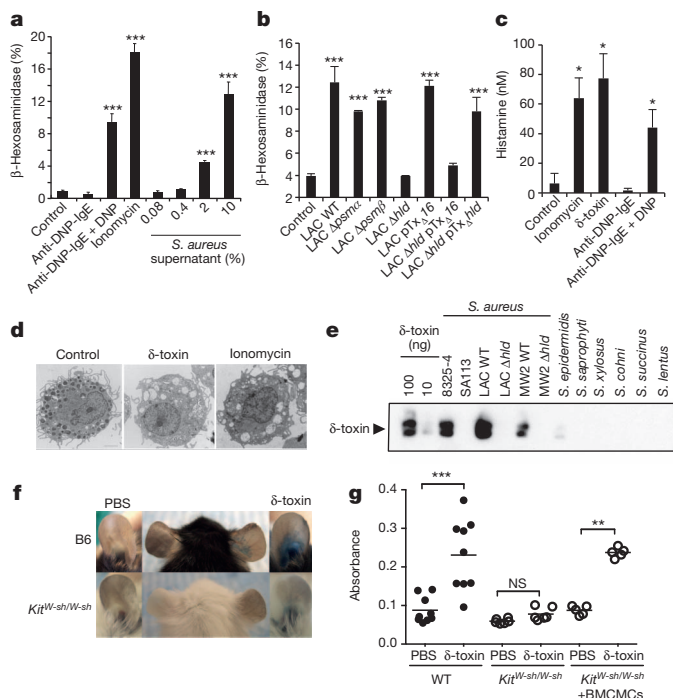


Figure 1 | *S. aureus* δ -toxin induces MC degranulation *in vitro* and *in vivo*.

a, Activity of β -hexosaminidase released to the extracellular media of BMCMCs stimulated with medium alone (control) or indicated stimuli including different concentrations of culture supernatant of *S. aureus* 8325-4. **b**, Activity of β -hexosaminidase in supernatants of MC/9 cells stimulated with 10% of culture supernatant from LAC *S. aureus* wild type (LAC WT) or isogenic mutants deficient in PSM- α peptides (LAC Δ psm α), PSM- β peptides (LAC Δ psm β), δ -toxin (LAC Δ hld), LAC wild type expressing vector alone (LAC pTX Δ 16), LAC deficient in δ -toxin expressing vector alone (LAC Δ hld pTX Δ 16) and strain complemented with δ -toxin plasmid (LAC Δ hld pTX Δ hld). Control represents 10% tryptic soy broth (TSB) medium. **c**, Histamine concentrations in culture supernatant of FSMCs after stimulation with indicated stimuli including synthetic δ -toxin at 30 μ g ml $^{-1}$ for 15 min. Data represent means \pm s.d. of triplicate cultures. Results are representative of at least three independent experiments (**a–c**). **d**, Electromicroscopic images of FSMCs stimulated with synthetic δ -toxin (30 μ g ml $^{-1}$) for 15 min. Images of unstimulated (Control) and ionomycin-treated FSMCs are also shown. Representative of at least 20 images. **e**, Expression of δ -toxin in *Staphylococcus* culture supernatants (0.5 μ l per well). Loading of lanes with synthetic δ -toxin (10 ng, 100 ng) is shown as reference. Representative of three experiments. **f**, C57BL/6 (WT) and MC-deficient (*Kit*^{W-sh/W-sh}) mice were injected intradermally into the left and right ears with δ -toxin (100 μ g) or PBS, respectively. One representative mouse for each group is shown. Representative of eight mice per group. **g**, Quantification of Evans blue extracted from skin tissue of WT, *Kit*^{W-sh/W-sh} and *Kit*^{W-sh/W-sh} reconstituted with BMCMCs is shown. Dots represent individual ear samples from two independent experiments. NS, not significant; * P < 0.05; ** P < 0.01; *** P < 0.001, two-tailed t -test.

from *S. epidermidis*, a bacterium that is present in normal skin, possessed weak MC degranulation which correlated with smaller amounts of δ -toxin than that from *S. aureus* strains (Fig. 1e and Supplementary Fig. 6). Furthermore, deficiency of δ -toxin had a larger effect on MC degranulation in *S. aureus* than in *S. epidermidis* (Supplementary Fig. 6). To assess whether δ -toxin induces MC degranulation *in vivo*, we injected synthetic δ -toxin into the skin of mouse ears and monitored MC degranulation by the vascular leakage of Evan's blue dye into the extravascular space using the passive cutaneous anaphylaxis (PCA) assay. Intradermal administration of δ -toxin induced Evan's blue dye leaking at the site of injection in wild-type mice, but not in MC-deficient *Kit*^{W-sh/W-sh} mice (Fig. 1f, g). Reconstitution of the skin of *Kit*^{W-sh/W-sh} mice with bone-marrow-derived cultured MCs (BMCMCs) restored leaking of the dye upon administration of δ -toxin (Fig. 1g). Moreover, the culture

supernatant from the δ -toxin-positive LAC strain induced Evan's blue dye leaking whereas that from δ -toxin-negative LAC Δ hld and SA113 strains did not (Supplementary Fig. 7). These results indicate that δ -toxin induces MC degranulation *in vitro* and *in vivo*.

Ca^{2+} influx in human neutrophils is triggered by δ -toxin through FPR2 (ref. 11). Because Ca^{2+} influx is an essential step in MC degranulation, we analysed whether δ -toxin induces Ca^{2+} influx in MCs. Stimulation of MCs with ionomycin or dinitrophenol (DNP) plus anti-DNP immunoglobulin-E (IgE) induced rapid Ca^{2+} influx (Fig. 2a). Likewise, δ -toxin triggered Ca^{2+} influx and this was abrogated by treatment with the Ca^{2+} chelator ethylene glycol tetra-acetic acid (EGTA) (Fig. 2a). EGTA also blocked MC degranulation induced by ionomycin, DNP plus anti-DNP IgE or δ -toxin (Fig. 2b). Similarly, MC degranulation induced by DNP plus anti-DNP IgE or δ -toxin was inhibited by the phosphoinositide 3-kinase inhibitor LY294002 (Fig. 2c). However, unlike antigen plus IgE, MC degranulation induced by δ -toxin did not require spleen tyrosine kinase (Syk) (Fig. 2d). Fpr1, Fpr2 and related family members were expressed in mouse MCs although their expression was higher in neutrophils (Supplementary Fig. 8). WRW4, a peptide antagonist of formyl peptide receptor 2 (FPR2), blocks human and mouse neutrophil activation induced by PSMs including δ -toxin¹¹. Notably, WRW4 inhibited mouse MC degranulation induced by δ -toxin both *in vitro* and *in vivo* (Supplementary Fig. 9a, b). Cyclosporin H, an antagonist of human FPR1, also partly inhibited mouse MC degranulation induced by δ -toxin (Supplementary Fig. 9c). However, human FPR2 ligands, MMK1 and lipoxin A4 did not induce mouse MC degranulation (Supplementary Fig. 10a). Furthermore, treatment with pertussis toxin (PTX), an inhibitor of G-protein coupled receptors, partly reduced MC degranulation induced by δ -toxin (Supplementary Fig. 10b). However, MCs from wild-type and *Fpr2*^{-/-} mice showed comparable MC degranulation induced by δ -toxin (Supplementary Fig. 10c). Collectively, these results indicate that δ -toxin induces MC degranulation through a signalling pathway that is different from that induced through antigen and IgE.

Stimulation with IgE and antigen, but not monomeric IgE, induces robust MC degranulation⁴. Notably, pre-incubation of MCs with anti-DNP or anti-trinitrophenyl (TNP) IgE alone markedly increased the degranulation activity of δ -toxin (Fig. 3a). The synergistic effect of monomeric IgE and δ -toxin was abrogated in MCs deficient in Syk (Fig. 3b). To test whether the synergism between monomeric IgE and δ -toxin could be observed *in vivo*, we injected monomeric IgE and δ -toxin (at concentrations not inducing MC degranulation) into the skin of mice and monitored MC degranulation *in vivo* with the PCA assay. At these inactive concentrations, δ -toxin induced Evans blue dye leaking at the site of injection in mice pre-treated with anti-DNP (Fig. 3c). These results indicate that IgE increases the MC degranulation activity of δ -toxin in the absence of antigen.

RNAIII, a regulatory RNA that is induced by the *agr* quorum-sensing system of *S. aureus*, encodes δ -toxin¹². Notably, supernatants from 26 *S. aureus* strains isolated from the lesional skin of patients with atopic dermatitis produced δ -toxin (Supplementary Fig. 11a). Moreover, RNAIII expression was detected in lesional skin colonized with *S. aureus*, but not normal skin, of patients with atopic dermatitis (Supplementary Fig. 11b, c). To test whether δ -toxin plays a role in allergic skin disease, we used a modified epicutaneous disease model in which the skin of BALB/c mice was colonized with wild-type or δ -toxin-deficient *S. aureus* and then challenged once with ovalbumin (OVA) to assess antigen-specific IgE production (Fig. 4a). One week after colonization with wild-type *S. aureus*, the mice developed severely inflamed reddened skin at the site of application (Fig. 4b, c). Expression of *hld* was detected in the skin on day 4 after bacterial colonization using a bioluminescent reporter *S. aureus* strain (Supplementary Fig. 12). Histological analysis showed spongiosis, parakeratosis and marked neutrophil-rich inflammatory infiltrates in the skin of mice colonized with wild-type *S. aureus* (Fig. 4c, d). In contrast, mice colonized with *S. aureus* lacking δ -toxin showed a significantly reduced skin inflammatory cell infiltrate and

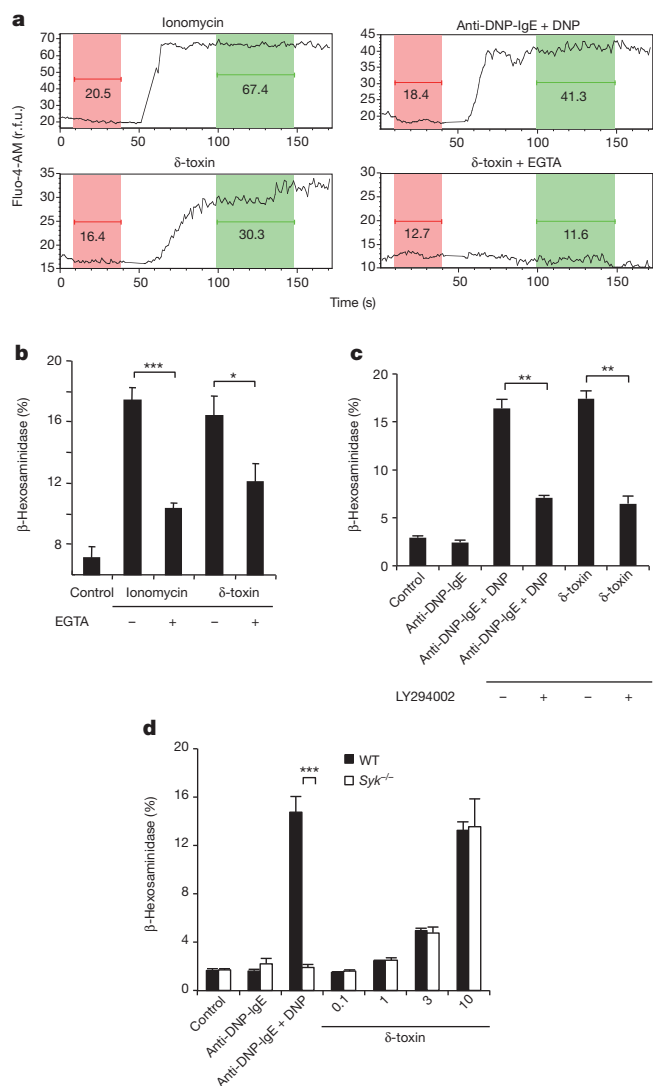


Figure 2 | MC degranulation induced by δ -toxin depends on Ca^{2+} influx/phosphoinositide 3-kinase pathway, but is independent of Syk. **a**, FSMCs loaded with the fluorescent Ca^{2+} indicator Fluo-4AM with or without EGTA were stimulated for 50 s. Baseline fluorescence (red) was measured, then the MCs were stimulated with indicated stimuli and fluorescence shift (green) was measured. RFU, relative fluorescence units. **b**, **c**, Activity of β -hexosaminidase in culture supernatants of FSMCs pre-treated with EGTA (**b**) or LY294002 (**c**) stimulated with medium alone (control), ionomycin, DNP-HSA (DNP) plus anti-DNP-IgE or δ -toxin ($10 \mu\text{g ml}^{-1}$). **d**, Activity of β -hexosaminidase in culture supernatants of FSMCs derived from $Syk^{-/-}$ and wild-type (WT) mice stimulated with the indicated concentration of δ -toxin (micrograms per millilitre). Data represent means \pm s.d. of triplicate cultures and are representative of at least three independent experiments (**b–d**). NS, not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$, two-tailed t -test.

disease score (Fig. 4b–d). Complementation of the *Ahld* mutant with a plasmid producing δ -toxin restored the disease scores comparable to those observed with the wild-type bacterium (Supplementary Fig. 13). The differential ability of wild-type and mutant *S. aureus* to promote inflammatory disease was not explained by differences in skin colonization (Supplementary Fig. 14a, b). Furthermore, mice colonized with wild-type *S. aureus* developed greater amounts of total serum IgE and IgG1, but not IgG2a, as well as IL-4 in the skin than mice inoculated with the δ -toxin mutant bacterium (Fig. 4e and Supplementary Figs 14c and 15). At 3 weeks, there was a slight increase in IgG1 production in mice colonized with the δ -toxin mutant bacterium compared with PBS control (Supplementary Fig. 15c), suggesting the existence of a minor pathway

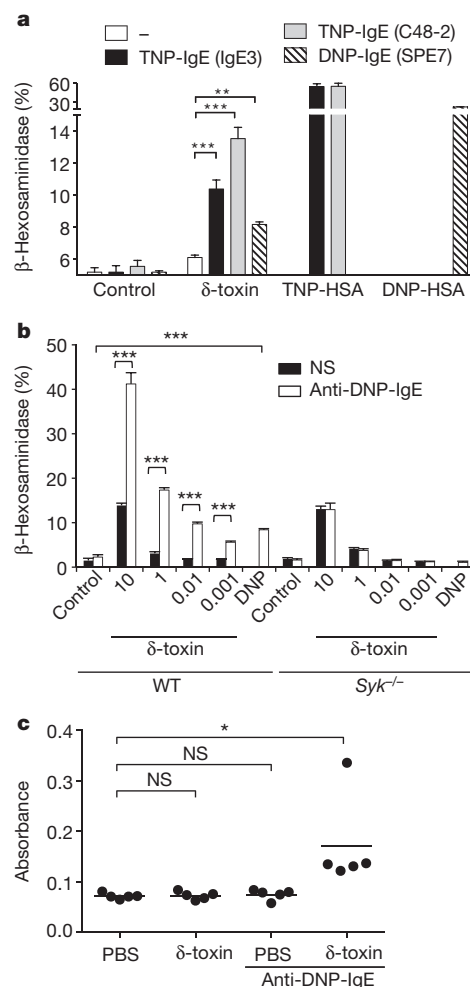


Figure 3 | Antigen-independent IgE signalling enhances δ -toxin-induced MC activation. **a**, Activity of β -hexosaminidase in culture supernatants of FSMCs stimulated with or without anti-DNP-IgE or TNP-IgE and then re-stimulated with δ -toxin ($0.01 \mu\text{g ml}^{-1}$), DNP-HSA (DNP) or TNP-HSA (TNP). **b**, Activity of β -hexosaminidase in culture supernatants of FSMCs derived from $Syk^{-/-}$ and wild-type mice (WT) pre-treated with or without anti-DNP-IgE, then stimulated with the indicated concentration of δ -toxin ($\mu\text{g ml}^{-1}$). Representative of at least three independent experiments. *** $P < 0.001$, two-tailed t -test (**a**, **b**). **c**, Quantification of Evans blue extracted from skin tissue of C57BL/6 mice injected intradermally into the left and right ears with δ -toxin ($5 \mu\text{g}$) or PBS, respectively. Data represent means \pm s.d. of triplicate cultures and are representative of at least three independent experiments (**a**, **b**). Dots represent individual ear samples. Representative of two independent experiments (**c**). NS, not significant; * $P < 0.05$, one-way analysis of variance with Tukey post-hoc test for multiple comparisons.

for IgG1 production dependent on *S. aureus* but independent of δ -toxin. In addition, pre-colonization with wild-type, but not the δ -toxin-deficient, *S. aureus* enhanced the production of OVA-specific IgE (Fig. 4f). Colonization with *S. aureus* without disrupting the skin barrier by stripping also induced inflammatory disease and enhanced IgE responses (Supplementary Fig. 16). Pre-colonization with δ -toxin-producing *S. aureus* was important to elicit antigen-specific IgE because administration of OVA before or concurrent with *S. aureus* colonization did not enhance OVA-specific IgE production (Supplementary Fig. 17). To test whether δ -toxin is sufficient to trigger allergic skin disease, we epicutaneously sensitized the skin of mice with OVA in the presence and absence of δ -toxin and challenged the mice with OVA alone or OVA plus δ -toxin 3 weeks later. We found that δ -toxin triggered inflammatory skin disease including OVA-specific IgE and IgG1 production whereas challenge with OVA alone did not (Supplementary Fig. 18). C57BL/6 mice colonized

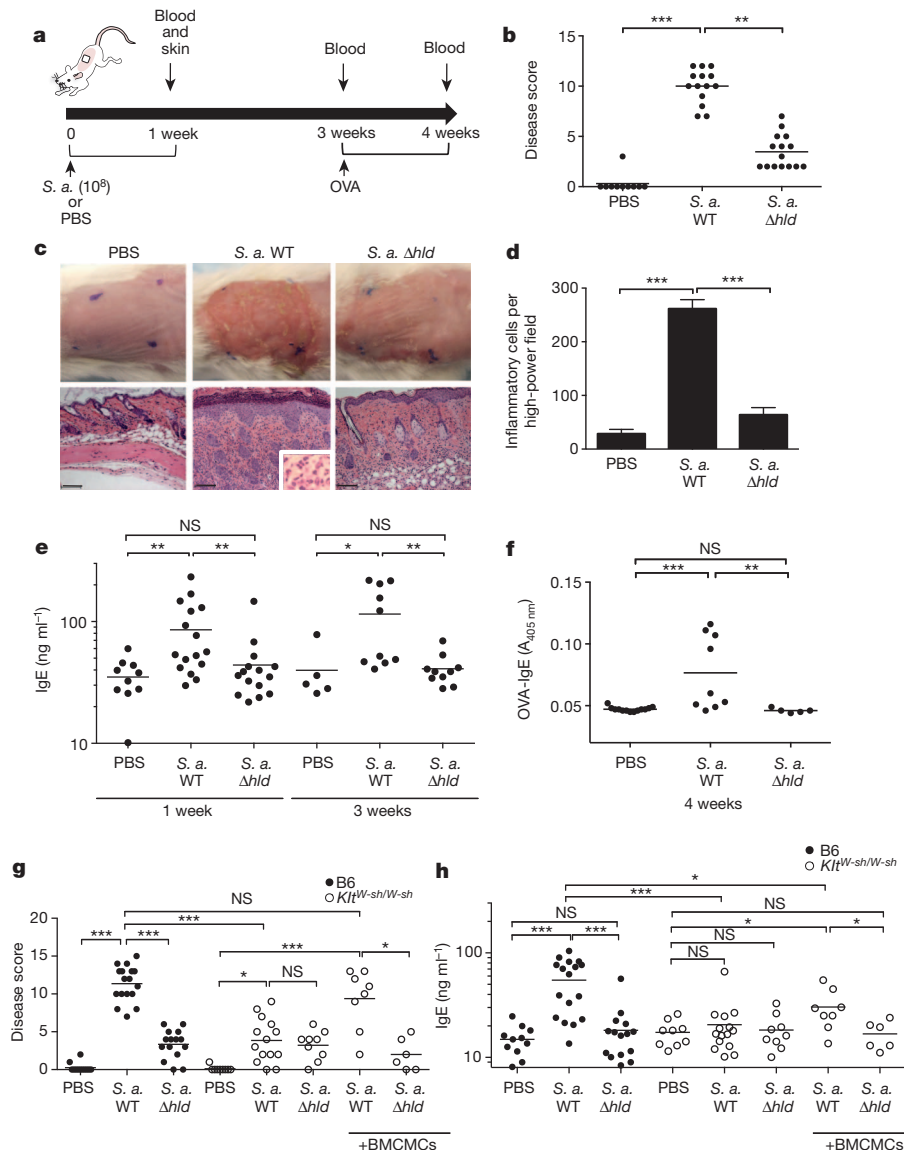


Figure 4 | *Staphylococcus* δ -toxin promotes IgE production and inflammatory skin disease by mast cells. **a**, *S. aureus* (*S. a.*) colonization and OVA sensitization protocol. Mice were colonized epicutaneously with 10^8 colony-forming units of *S. aureus* using a gauze patch for 1 week. For OVA sensitization, a patch containing OVA or PBS was applied to the same skin site 2 weeks after *S. aureus* inoculation. **b**, Skin disease score 1 week after colonization with wild-type and δ -toxin mutant (Δhld) *S. aureus* or treated with PBS. $**P < 0.01$; $***P < 0.001$, Kruskal–Wallis test with post-hoc Dunn’s test for multiple comparisons. **c**, Skin phenotype and histopathology of BALB/c mice colonized with *S. aureus* or treated with PBS. Skin sections were stained with haematoxylin and eosin. Scale bar, 100 μ m. Inset shows high-power image with neutrophil-rich inflammation. Representative of 14 mice per group. **d**, Number of inflammatory cells in skin of BALB/c mice colonized with *S. aureus* or treated with PBS. Results are depicted as the number of

inflammatory cells per high-power field. Error bars, means \pm s.e.m. **e**, Concentrations of serum IgE in BALB/c mice colonized with *S. aureus* or treated with PBS at 1 and 3 weeks after colonization with *S. aureus*. **f**, Concentrations of serum OVA-specific IgE after OVA sensitization in BALB/c mice colonized with *S. aureus* or treated with PBS. A405, absorbance at 405 nm. **g**, Skin disease score in C57BL/6 (B6), MC-deficient (*Kit*^{W-sh/W-sh}) and MC-deficient (*Kit*^{W-sh/W-sh}) mice reconstituted with MCs at 1 week after the inoculation with *S. aureus*. **h**, Concentrations of serum IgE 1 week after colonization of B6, *Kit*^{W-sh/W-sh} and *Kit*^{W-sh/W-sh} mice reconstituted with MCs with wild-type and δ -toxin mutant (Δhld) *S. aureus* or treated with PBS. Dots represent individual mice pooled from two independent experiments. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$, one-way analysis of variance with Tukey post-hoc test for multiple comparisons (**e**–**h**).

with wild-type *S. aureus* also developed higher concentrations of serum IgE and more severe inflammatory skin disease than mice inoculated with the bacterium deficient in δ -toxin (Fig. 4g, h). MC-deficient *Kit*^{W-sh/W-sh} mice inoculated with wild-type *S. aureus* showed reduced concentrations of IgE serum and skin inflammation than wild-type mice (Fig. 4g, h). Adoptive transfer of MCs into the skin of *Kit*^{W-sh/W-sh} mice restored skin disease and increased IgE production in mice colonized with wild type, but not *S. aureus* lacking δ -toxin (Fig. 4g, h and Supplementary Fig. 19). There were increased numbers of *S. aureus* and total bacteria in the skin of *Kit*^{W-sh/W-sh} mice (Supplementary Fig. 19), suggesting that mast cells can regulate bacterial colonization under our

experimental conditions. Microscopic analysis showed that the dermal MC densities in the skin of *Kit*^{W-sh/W-sh} recipient mice were approximately 50% of those found in age-matched C57BL/6 mice (Supplementary Fig. 19). Furthermore, toluidine-positive granules associated with MC degranulation were present in the skin of mice colonized with wild-type, but not δ -toxin-deficient, *S. aureus* (Supplementary Fig. 19). Taken together, these results indicate that δ -toxin from *S. aureus* promotes allergic skin disease through activation of MCs.

The δ -toxin transcript is contained in RNAIII, a regulatory RNA that governs *S. aureus* virulence genes^{13,14}. The role of δ -toxin in the growth of *S. aureus* is not understood. Because δ -toxin can form pores

on the surface of certain bacteria¹⁵, one possibility is that it promotes pathogen colonization by killing competing bacteria. Our results indicate that the host senses *S. aureus* through the detection of δ -toxin to promote innate and adaptive Th2 immune responses by MC degranulation. Although clinical studies are needed to determine the role of δ -toxin in atopic dermatitis, our results in mouse models suggest that in the setting of genetic defects associated with the disease², δ -toxin may promote allergic immune responses and that strategies to inhibit δ -toxin might be beneficial for the treatment of atopic dermatitis.

METHODS SUMMARY

Culture of mast cells and degranulation. Preparations of BMCs and fetal skin-derived mast cells (FSMCs) were previously described¹⁶. The purity of MCs was greater than 95% as assessed by surface expression of Fc ϵ RI and CD117 (eBioscience). Degranulation of MCs was assessed by β -hexosaminidase assay as described¹⁶.

PCA assay. PCA assay was performed as described with minor modifications¹⁷.

Epicutaneous sensitization with *S. aureus*. The dorsal skin of 6- to 8-week-old female mice was shaved and stripped using a transparent bio-occlusive dressing (Tegaderm; 3M). One hundred million colony-forming units of *S. aureus* strains were placed on a patch of sterile gauze and attached to the shaved skin with another transparent bio-occlusive dressing (Tegaderm; 3M). Each mouse was exposed to *S. aureus* for 1 week through the patch. After a 2 week interval, each mouse was challenged once with 100 μ g ovalbumin epicutaneously for 1 week and the animals then killed for analyses.

Animal study. All animal studies were performed according to approved protocols by the University of Michigan Committee on the Use and Care of Animals.

Statistical analysis. All analyses were performed using GraphPad Prism. Differences were considered significant when $P < 0.05$.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 November 2012; accepted 12 September 2013.

Published online 30 October 2013.

- Williams, H. & Flohr, C. How epidemiology has challenged 3 prevailing concepts about atopic dermatitis. *J. Allergy Clin. Immunol.* **118**, 209–213 (2006).
- Elias, P. M. & Steinhoff, M. “Outside-to-inside” (and now back to “outside”) pathogenic mechanisms in atopic dermatitis. *J. Invest. Dermatol.* **128**, 1067–1070 (2008).
- Liu, F. T., Goodarzi, H. & Chen, H. Y. IgE, mast cells, and eosinophils in atopic dermatitis. *Clin. Rev. Allergy Immunol.* **41**, 298–310 (2011).
- Galli, S. J. & Tsai, M. IgE and mast cells in allergic disease. *Nature Med.* **18**, 693–704 (2012).
- Rudikoff, D. & Lebowitz, M. Atopic dermatitis. *Lancet* **351**, 1715–1721 (1998).
- Leung, D. Y., Walsh, P., Giorno, R. & Norris, D. A. A potential role for superantigens in the pathogenesis of psoriasis. *J. Invest. Dermatol.* **100**, 225–228 (1993).
- Supajatura, V. et al. Differential responses of mast cell Toll-like receptors 2 and 4 in allergy and innate immunity. *J. Clin. Invest.* **109**, 1351–1359 (2002).
- Selander, C., Engblom, C., Nilsson, G., Scheynius, A. & Andersson, C. L. TLR2/MyD88-dependent and -independent activation of mast cell IgE responses by the skin commensal yeast *Malassezia sympodialis*. *J. Immunol.* **182**, 4208–4216 (2009).
- Schmalzer, M. et al. Lipoproteins in *Staphylococcus aureus* mediate inflammation by TLR2 and iron-dependent growth *in vivo*. *J. Immunol.* **182**, 7110–7118 (2009).
- Wang, R. et al. Identification of novel cytolytic peptides as key virulence determinants for community-associated MRSA. *Nature Med.* **13**, 1510–1514 (2007).
- Kretschmer, D. et al. Human formyl peptide receptor 2 senses highly pathogenic *Staphylococcus aureus*. *Cell Host Microbe* **7**, 463–473 (2010).
- Novick, R. P. Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol. Microbiol.* **48**, 1429–1449 (2003).
- Morfeldt, E., Tegmark, K. & Arvidson, S. Transcriptional control of the agr-dependent virulence gene regulator, RNAIII, in *Staphylococcus aureus*. *Mol. Microbiol.* **21**, 1227–1237 (1996).
- Sugiyama, Y. et al. Changes in the agr locus affect enteritis caused by methicillin-resistant *Staphylococcus aureus*. *J. Clin. Microbiol.* **47**, 1528–1535 (2009).
- Cogen, A. L. et al. Selective antimicrobial action is provided by phenol-soluble modulins derived from *Staphylococcus epidermidis*, a normal resident of the skin. *J. Invest. Dermatol.* **130**, 192–200 (2010).
- Yamada, N., Matsushima, H., Tagaya, Y., Shimada, S. & Katz, S. I. Generation of a large number of connective tissue type mast cells by culture of murine fetal skin cells. *J. Invest. Dermatol.* **121**, 1425–1432 (2003).
- Wershil, B. K., Wang, Z. S., Gordon, J. R. & Galli, S. J. Recruitment of neutrophils during IgE-dependent cutaneous late phase reactions in the mouse is mast cell-dependent. Partial inhibition of the reaction with antiserum against tumor necrosis factor- α . *J. Clin. Invest.* **87**, 446–453 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Koonse for animal husbandry, J. Whitfield for enzyme-linked immunosorbent assays, S. Meshinchi for electron microscopy, V. Basrur for mass spectrometry, K. Kidwell for advice with statistical analysis, M. K. Oyoshi and R. S. Geha for experimental advice, V. Y. Tan for help with constructing the LAC P3-*lux* strain and A. Burberry for reviewing the manuscript. Y.N. was supported by fellowships from the Chiba University Global COE Program, the Cell Science Research Foundation and the Kanoe Foundation for the Promotion of Medical Science. J.O. and K.B.C. were supported by Department of Veterans Affairs Merit Award IO1BX000429. A.E.V., G.Y.C.C. and M.O. were supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH). This work supported by NIH grants R01AR059688 to G.N. and R01HL062996 to J.B.T. and funds to the Michigan Comprehensive Cancer Center Immunology Monitoring Core from the University of Michigan's Cancer Center Support Grant.

Author Contributions Y.N., N.I. and G.N. designed the research. Y.N. conducted the experiments and analysed data with the help of R.M.-P., S.M.C. and M.H. J.O., K.B.C., J.B.T. and M.J.M. generated and provided critical reagents or material. A.E.V., G.Y.C.C. and M.O. engineered bacterial strains. Y.N. and G.N. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.N. (gabriel.nunez@umich.edu).

METHODS

Bacterial strains. *S. aureus* strain 8325-4 and its isogenic toxin mutant ($\Delta\alpha\beta\gamma$) have been previously described¹⁸. *S. aureus* strains SA113 and Newman, and isogenic mutants deficient in lipoprotein diacylglycerol transferase (*AlgT*), have also been previously described¹⁹. *S. aureus* strains LAC and MW2, their isogenic δ -toxin mutants (Δhld), the *psm* gene deleted mutants ($\Delta psmx$, $\Delta psmb$) and LAC *agr* mutant (Δagr) have been previously described¹⁰. The isogenic Δhld mutant of *S. epidermidis* 1457, a clinical isolate²⁰, was produced by an allelic replacement procedure²¹. This was done in a way analogous to the *S. aureus* Δhld mutants used herein, abolishing translation by exchanging the third base in the *hld* start codon from ATG to ATA (to avoid interfering with the function of RNAPIII). LAC P3-*lux* was constructed by integration of the *S. aureus* LAC *agr* P3 promoter fused to the *luxABCDE* operon of *Photobacterium luminescens* with codon use optimized for staphylococci²² into the $\Phi 11$ attB site of the *S. aureus* genome, using a procedure described by Luong and Lee²³. Plasmid pTX Δhld was constructed by cloning the *hld* coding sequence containing the ribosomal binding site region in the BamHI/MluI sites of plasmid pTX Δ ¹⁰. The *hld* gene was amplified from the genomic DNA of the respective strain, because the δ -toxin sequence differs in one amino acid in position 10 (serine or glycine) in these two strains. The δ -toxin is constitutively expressed in these plasmids. See Supplementary Table 1 for all oligonucleotides used in generation of the strains. Clinical isolates of *S. aureus* from children diagnosed with atopic dermatitis were obtained originally from the Department of Laboratory Medicine and Pathobiology at the University of Toronto²⁴. *S. epidermidis* (NI335), *Staphylococcus cohnii* (NI446), *S. saprophyticus* (NI488), *Staphylococcus xylosus* (NI987), *Staphylococcus sciuri* (NI981), *Staphylococcus succinus* (NI534), *Staphylococcus lentus* (NI487) and *Staphylococcus fleuretti* (NI533) were isolated by plating on BHI after culturing at 37 °C for 2 days under aerobic conditions. Identification of bacterial species was verified by 16S rRNA gene sequencing as described²⁵. Bacterial supernatants were produced by overnight culture with shaking in tryptic soy broth (TSB) followed by filtration through a 0.2 μ m filter.

Mice. C57BL/6, C57BL/6-Kit^{W-sh/Kit^{W-sh}} (B6.CG-Kit^{W-sh}/HNhrJaeBsmJ) and BALB/c mice were purchased from Jackson Laboratories. Syk^{+/-} mouse breeders were a gift from S. Teitelbaum and Syk^{-/-} embryos were generated by intercrossing. We used 4- to 12-week-old age-matched female mice for *in vivo* experiments. Mice were allocated randomly into experimental groups. All mouse strains were housed under pathogen-free conditions. The animal studies were conducted under approved protocols by the University of Michigan Committee on Use and Care of Animals.

Materials. The synthetic peptides fPSM- $\alpha 2$ (fMGIIAGIIKVIKSLIEQFTGK), fPSM- $\alpha 3$ (fMGIIAGIIKFIKGLIEKFTGK), f δ -toxin (fMAQDIISTIGDLVKWIIDTVNKFTEK), (WRWWW-CONH2) and MMK-1 (LESIFRSLFRVM) were purchased from American Peptide. Unformylated δ -toxin (MAQDIISTIGDLVKWIIDTVNKFTEK) was synthesized at The University of Michigan Protein Structure Facility. Polyclonal anti- δ -toxin antibody was produced in rabbits by immunization with a synthetic multiple antigenic peptide showing an 18 amino-acid peptide (IGDLVKWIIDTVNKFTEK) (Sigma-Genosys) from the full-length δ -toxin sequence. Rabbit IgG was purified from rabbit serum on Protein A (Pierce) according to the manufacturer's protocol.

Protein purification from *S. aureus* culture supernatant. *S. aureus* was cultured in 700 ml chemical defined medium supplemented with 2% yeast extract²⁶. Filtrate cultured supernatant was incubated with carboxymethyl cellulose equilibrated with 10 mM sodium citrate (pH 5.5), and eluted with a linear gradient of 0–1 M NaCl. Fractions containing β -hexosaminidase activity were collected and adjusted at pH 7.4, 100 mM HEPES. The sample was concentrated using Amicon Ultra-15, 5 kDa filter (Millipore). Concentrated sample was further fractionated with a Superdex 200 10/300 GL column (GE). Final positive fractions were pooled and concentrated using an Amicon Ultra-15 filter (Supplementary Fig. 2b).

Protein identification by liquid chromatography–tandem mass spectrometry (LC–MS/MS). Purified sample was denatured in 8 M urea, reduced by incubation with 10 mM DTT at 37 °C for 30 min and alkylated using 50 mM iodoacetamide at room temperature for 30 min. The protein sample was digested with sequencing grade trypsin (Promega) overnight at 37 °C. The reaction was terminated by acidification with trifluoroacetic acid (0.1% v/v) and peptides were purified using a SepPak C18 cartridge following the manufacturer's protocol (Waters Corporation). Eluted peptides were directly introduced into an ion-trap mass spectrometer (LTQ-XL, Thermo Fisher) equipped with a nano-spray source. The mass spectrometer was operated in data-dependent MS/MS mode to acquire a full MS scan (400–2000 *m/z*) followed by MS/MS on the top six ions from the full MS scan. Dynamic exclusion was set to collect two MS/MS spectra on each ion and exclude it for a further 2 min. Raw files were converted to mzXML format and searched against the *S. aureus* NCTC 8325 database appended with decoy (reverse) database using X! Tandem with k-score plug-in, an open-source search engine developed by the Global Proteome Machine (<http://www.thegpm.org>). Search parameters included a precursor

peptide mass tolerance window of 1 Da and fragment mass tolerance of 0.5 Da. Oxidation of methionine (+16 Da), and carbamidomethylation of cysteines (+57 Da) were considered as variable modifications. The search was restricted to tryptic peptides with one missed cleavage. Results of the X! Tandem search were then subjected to Trans-Proteomic Pipeline (TPP) analysis, a suite of software including PeptideProphet and ProteinProphet. All proteins with a ProteinProphet probability of greater than 0.9 were considered positive and verified manually.

Culture of mast cells and degranulation. Preparations of BMCMCs and fetal skin-derived mast cells (FSMCs) were previously described¹⁶. Bone marrow cells from *Fpr2*^{-/-} mice were provided by J. M. Wang. The purity of MCs was greater than 95% as determined by surface expression of Fc ϵ RI and CD117 (eBioscience). Degranulation of MCs was assessed by β -hexosaminidase assay as previously described¹⁶. Briefly, MCs (2×10^6 ml⁻¹) were preloaded with or without IgEs (anti-DNP IgE (clone; SPE7); 0.3 μ g ml⁻¹, anti-TNP IgE (clone; IgE3 and C48-2); 0.5 μ g ml⁻¹) in RPMI with IL-3 for 15 h. The cells were re-suspended in Tyrode's buffer (Sigma) at 2×10^4 cells per 100 μ l for FSMCs or 1×10^5 cells per 100 μ l for BMCMCs and MC/9 cells, aliquoted in triplicate into a 96-well U-bottom plate and incubated with EGTA (1 mM, Sigma), LY294002 (100 μ M, Sigma), WRW4 (10 μ M) and Cyclosporine H (10 μ M, Alexis Biochemicals) for 30 min, then stimulated with DNP-HSA (30 ng ml⁻¹), TNP-HSA (30 nM) for 30 min, ionomycin (1 μ M, Sigma), δ -toxin (indicated concentrations), PSM- α s (indicated concentrations) or FPR2 ligands for 15 min. Results of various stimuli are given as a relative percentage, where freeze and thaw of total cell culture represents 100%.

MC reconstitution in Kit^{W-sh/Kit^{W-sh}} mice. For BMCMC reconstitution experiments, 10^6 BMCMCs (cell purity was greater than 95%) were injected into the ear skin. Four million BMCMCs in 50 μ l \times eight injections were injected into the shaved back skin of non-randomized Kit^{W-sh/Kit^{W-sh}} mice as described²⁷. Four to six weeks later, the mice were subjected to experimental PCA assay or epicutaneous *S. aureus* sensitization. The number of animals per group ($n = 5$ –8) was chosen as the minimum probably required for conclusions of biological significance, established from previous experience. The reconstitution rate of cutaneous MCs was quantified blindly by an independent observer and scored as the number of MCs per low-power field in toluidine blue stained tissue slides by microscope. The average rate of reconstituted MCs was approximately 40% in the ear pinna and 50% in the back skin (Supplementary Figs 19 and 20).

PCA assay. PCA assay was performed as previously described with minor modifications¹⁷. Ears of non-randomized mice were injected intradermally with or without α -DNP-IgE in 40 μ l saline; 15 h later, mice were challenged with 20 μ l saline with or without synthetic δ -toxin (100 μ g or 5 μ g) or TSB bacteria supernatants. The number of animals per group ($n = 5$ –8) was chosen on the basis of previous experience as the minimum probably required for conclusions of biological significance. After inoculation, 0.1 ml of 5 mg ml⁻¹ Evans blue dye was injected intravenously. Extravasation of Evans blue dye was monitored for 30 min, and 4 mm of punched-out biopsies were incubated at 63 °C overnight in 200 μ l formamide. Quantitative analysis of extracts was determined by measuring the absorbance at 600 nm.

Ca²⁺ influx assay. FSMCs (2×10^6 ml⁻¹) were preloaded with or without anti-DNP-IgE (0.3 μ g ml⁻¹) in RPMI with IL-3 for 15 h. Cells were washed and then loaded with Fluo-4AM (5 μ M, Life Technologies) for 30 min. Cells were washed again and further incubated in Tyrode's buffer with or without EGTA (1 mM) for 30 min. DNP-HSA (30 ng ml⁻¹), ionomycin (1 μ M) or δ -toxin (30 μ g ml⁻¹) were used to induce calcium flux in these cells. Ca²⁺ flux was measured using a flow cytometer (FACSCalibur, BD Biosciences) to monitor relative fluorescence units (RFU) as described²⁸.

Epicutaneous sensitization with *S. aureus* or OVA. We performed epicutaneous colonization with *S. aureus* by shaving the dorsal skin of non-randomized 6- to 8-week-old female mice and three-time stripping using a transparent bio-occlusive dressing (Tegaderm; 3M). Sample size ($n = 5$ –8 per group) was based on previous experience as the size necessary for conclusions of biological significance and adequate statistical analysis. After overnight culture at 37 °C with shaking, *S. aureus* were cultured in fresh TSB medium for 4 h at 37 °C with shaking, washed and re-suspended in PBS at 10^8 colony-forming units of *S. aureus* LAC or LAC (Δhld) strains. One hundred microlitres of the *S. aureus* suspension was placed on a patch of sterile gauze (1 cm \times 1 cm) and attached to the shaved skin with transparent bio-occlusive dressing. Each mouse was exposed to *S. aureus* for 1 week through the patch. After a 2 week interval, each mouse was challenged once with 100 μ g OVA (Grade V, Sigma) epicutaneously for 1 week and the animals were killed for analyses. For OVA sensitization model, BALB/c mice were sensitized epicutaneously with OVA (100 μ g) with or without synthetic δ -toxin (100 μ g) for 1 week. After a 2 week interval, mice were challenged with OVA (100 μ g) with or without synthetic δ -toxin (100 μ g) at the same skin site.

Skin disease score. The severity of skin lesions was scored according to defined macroscopic diagnostic criteria in a blind fashion²⁹. In brief, the total clinical score

of skin lesions was designated as the sum of individual scores, graded as 0 (none), 1 (mild), 2 (moderate) and 3 (severe) for thickness, erythema, oedema, erosion and scaling.

Histology. Skin tissue was formalin fixed, paraffin embedded and sectioned for haematoxylin and eosin and toluidine blue staining.

Cytokine and immunoglobulin concentrations. Chemokines and cytokines were measured with enzyme-linked immunoabsorbent assay (ELISA) kits (R&D Systems). For tissue cytokines, skin tissue (5 mm × 10 mm area) was removed and homogenized. The skin homogenates were centrifuged and supernatants were collected for cytokine measurements by ELISA. Serum IgG1 and IgG2a were measured with ELISA kit (Cayman chemical). Serum IgE was measured with ELISA kit (Bethyl Laboratories). ELISA for OVA-IgE was described previously³⁰.

RNA isolation from human skin samples. Wash fluid derived from lesional and normal skin of patients with atopic dermatitis was collected using a 2.5-cm-diameter polypropylene chamber as reported³¹. One hundred microlitres of the samples were mixed with an equal volume of RNeasy Protect Bacteria Reagent (QIAGEN) and RNA extracted with a Bacterial RNA Kit (OMEGA). The human studies were approved by the Indiana University Institutional Review Committee³¹. Informed consent was obtained from all participants.

Quantitative real-time PCR with reverse transcription. Complementary DNA was synthesized using a High Capacity RNA-to-cDNA Kit (Applied Biosystems), according to the manufacturer's instructions. Quantitative real time RT-PCR (qPCR) was performed using a SYBR green PCR master mix (Applied Biosystems) and StepOne Real-time PCR system (Applied Biosystems). Primers to amplify mouse *Fpr* genes³² and bacterial genes (*RNAIII*, *gyrB*, *16S rRNA*) have been described^{33,34}. Expression of mouse *Fpr* genes was normalized to that of *Gapdh* (F; 5'-CCTCGTCCCGTAGACAAAATG-3, R; 5'-TCTCCACTTTGCCACCTGCAA-3) and expression was analysed by the $2^{-\Delta\Delta Ct}$ method. *RNAIII* expression in human skin samples was normalized to that of *S. aureus gyrB* and that of *gyrB* to universal bacterial *16S rRNA*, and relative expression calculated by the $2^{-\Delta Ct}$ method. *RNAIII* and *gyrB* expression in some human skin samples was below the detection limit and arbitrarily given a value of zero for statistical analysis. LAC wild type and LAC *Δagr* cultured for 24 h were used as reference controls.

Measurement of P3-lx expression. To determine the amounts of P3-lx expression in culture, 10^5 ml⁻¹ LAC P3-lx strain was suspended in TSB and luminescence emitted from P3-lx-expressing bacteria was measured using a LMax luminometer (Molecular Devices). For *in vivo* bioluminescence imaging, mice were killed, the skin dressing removed and immediately placed into the light-tight chamber of the CCD (charge-coupled device) camera system (IVIS200, Xenogen). Luminescence emitted from lux-expressing bacteria in the tissue was quantified using the software program Living Image (Xenogen).

18. Nilsson, I. M., Hartford, O., Foster, T. & Tarkowski, A. Alpha-toxin and gamma-toxin jointly promote *Staphylococcus aureus* virulence in murine septic arthritis. *Infect. Immun.* **67**, 1045–1049 (1999).
19. Stoll, H., Dengiel, J., Nerz, C. & Gotz, F. *Staphylococcus aureus* deficient in lipidation of prelipoproteins is attenuated in growth and immune activation. *Infect. Immun.* **73**, 2411–2423 (2005).
20. Mack, D. *et al.* Characterization of transposon mutants of biofilm-producing *Staphylococcus epidermidis* impaired in the accumulative phase of biofilm production: genetic identification of a hexosamine-containing polysaccharide intercellular adhesin. *Infect. Immun.* **62**, 3244–3253 (1994).
21. Francis, K. P. *et al.* Monitoring bioluminescent *Staphylococcus aureus* infections in living mice using a novel luxABCDE construct. *Infect. Immun.* **68**, 3594–3600 (2000).
22. Bae, T. & Schneewind, O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid* **55**, 58–63 (2006).
23. Luong, T. T. & Lee, C. Y. Improved single-copy integration vectors for *Staphylococcus aureus*. *J. Microbiol. Methods* **70**, 186–190 (2007).
24. Yeung, M. *et al.* Identification of major clonal complexes and toxin producing strains among *Staphylococcus aureus* associated with atopic dermatitis. *Microbes Infect.* **13**, 189–197 (2011).
25. Hasegawa, M. *et al.* Differential release and distribution of Nod1 and Nod2 immunostimulatory molecules among bacterial species and environments. *J. Biol. Chem.* **281**, 29054–29063 (2006).
26. Miller, R. D. & Fung, D. Y. Amino acid requirements for the production of enterotoxin B by *Staphylococcus aureus* S-6 in a chemically defined medium. *Appl. Microbiol.* **25**, 800–806 (1973).
27. Grimbaldston, M. A. *et al.* Mast cell-deficient W-sh/c-kit mutant Kit W-sh/W-sh mice as a model for investigating mast cell biology in vivo. *Am. J. Pathol.* **167**, 835–848 (2005).
28. Vig, M. *et al.* Defective mast cell effector functions in mice lacking the CRACM1 pore subunit of store-operated calcium release-activated calcium channels. *Nature Immunol.* **9**, 89–96 (2008).
29. Leung, D. Y. *et al.* Thymopentin therapy reduces the clinical severity of atopic dermatitis. *J. Allergy Clin. Immunol.* **85**, 927–933 (1990).
30. Nakajima, S. *et al.* Langerhans cells are critical in epicutaneous sensitization with protein antigen via thymic stromal lymphopoietin receptor signaling. *J. Allergy Clin. Immunol.* **129**, 1048–1055 (2012).
31. Travers, J. B. *et al.* Infected atopic dermatitis lesions contain pharmacologic amounts of lipoteichoic acid. *J. Allergy Clin. Immunol.* **125**, 146–152 e141–142 (2010).
32. Riviere, S., Challet, L., Fluegge, D., Spehr, M. & Rodriguez, I. Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. *Nature* **459**, 574–577 (2009).
33. Seidl, K. *et al.* Relationship of *agr* expression and function with virulence and vancomycin treatment outcomes in experimental endocarditis due to methicillin-resistant *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* **55**, 5631–5639 (2011).
34. Barman, M. *et al.* Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infect. Immun.* **76**, 907–915 (2008).

HIV-1 evades innate immune recognition through specific cofactor recruitment

Jane Rasaiyaah¹, Choon Ping Tan¹, Adam J. Fletcher¹, Amanda J. Price², Caroline Blondeau¹, Laura Hilditch¹, David A. Jacques², David L. Selwood³, Leo C. James², Mahdad Noursadeghi^{1*} & Greg J. Towers^{1*}

Human immunodeficiency virus (HIV)-1 is able to replicate in primary human macrophages without stimulating innate immunity despite reverse transcription of genomic RNA into double-stranded DNA, an activity that might be expected to trigger innate pattern recognition receptors. We reasoned that if correctly orchestrated HIV-1 uncoating and nuclear entry is important for evasion of innate sensors then manipulation of specific interactions between HIV-1 capsid and host factors that putatively regulate these processes should trigger pattern recognition receptors and stimulate type 1 interferon (IFN) secretion. Here we show that HIV-1 capsid mutants N74D and P90A, which are impaired for interaction with cofactors cleavage and polyadenylation specificity factor subunit 6 (CPSF6) and cyclophilins (Nup358 and CypA), respectively^{1,2}, cannot replicate in primary human monocyte-derived macrophages because they trigger innate sensors leading to nuclear translocation of NF- κ B and IRF3, the production of soluble type 1 IFN and induction of an antiviral state. Depletion of CPSF6 with short hairpin RNA expression allows wild-type virus to trigger innate sensors and IFN production. In each case, suppressed replication is rescued by IFN-receptor blockade, demonstrating a role for IFN in restriction. IFN production is dependent on viral reverse transcription but not integration, indicating that a viral reverse transcription product comprises the HIV-1 pathogen-associated molecular pattern. Finally, we show that we can pharmacologically induce wild-type HIV-1 infection to stimulate IFN secretion and an antiviral state using a non-immunosuppressive cyclosporine analogue. We conclude that HIV-1 has evolved to use CPSF6 and cyclophilins to cloak its replication, allowing evasion of innate immune sensors and induction of a cell-autonomous innate immune response in primary human macrophages.

HIV-1 capsid (CA) mutant N74D cannot recruit CPSF6 and is insensitive to depletion of HIV-1 cofactors Nup358 and TNPO3, suggesting that it may use alternate cofactors for nuclear entry^{1–3}. Furthermore, unlike wild-type (WT) HIV-1, HIV-1 N74D cannot replicate in monocyte-derived macrophages (MDM) (Fig. 1a and Extended Data Fig. 2)^{2,4}. Remarkably, an inability to replicate was accompanied by a burst of IFN- β detected 2–5 days after low-multiplicity infection (Fig. 1b and Extended Data Fig. 2). The antiviral activity of IFN- β (Extended Data Fig. 3a)⁵ was revealed by rescuing HIV-1 N74D, but not WT replication with antibody to the IFN- α/β receptor α chain (IFNAR2) (Fig. 1c, d and Extended Data Fig. 3b). Co-infection of MDM with WT and HIV-1 N74D led to suppression of WT replication (Fig. 1e), which was also rescued by IFNAR2 antibody (Extended Data Fig. 3c). This demonstrated that sensitivity to IFN-mediated restriction was not limited to the mutant virus.

In contrast to the spreading infection assay in which HIV-1 N74D was completely suppressed, assessment of single-round infection in MDM with higher dose viral inocula revealed only a 5-fold reduction

of HIV-1 N74D infectivity compared to WT (Fig. 1f). However, this reduction was also restored to WT levels by IFNAR2 blockade (Extended Data Fig. 3d). In this experiment we did not detect IFN- β , probably owing to assay sensitivity, but interferon-stimulated genes (ISGs) *IP10* (also known as *CXCL10*), *IFIT1* and *CCL8*, were induced following infection with HIV-1 N74D, but not WT virus (Extended Data Fig. 4a). ISG induction was confirmed by microarray transcriptional profiling of host responses to HIV-1 N74D, which showed expected enrichment for innate immune type-1 IFN pathways at a genome-wide level (Extended Data Fig. 4c, d). These findings, together with the time course of IFN release during spreading infection (Fig. 1b), indicate that multiple-round replication amplifies virus-induced innate responses, leading to high levels of IFN- β secretion and potent suppression of HIV-1 replication (Fig. 1a).

We next measured ISG induction in HIV-1 N74D-infected MDM when either DNA synthesis or integration were prevented by mutation of reverse transcriptase⁶ (RT) or integrase⁷ (IN), respectively. Infection by HIV-1 CA(N74D), RT(D185E) double mutant did not stimulate *IP10* expression, whereas infection with HIV-1 double-mutant CA(N74D), IN(D116N) induced *IP10* expression comparable to the WT virus (Fig. 1g). These data indicate that the innate immune response in MDM depends on detection of the products of reverse transcription, not integration.

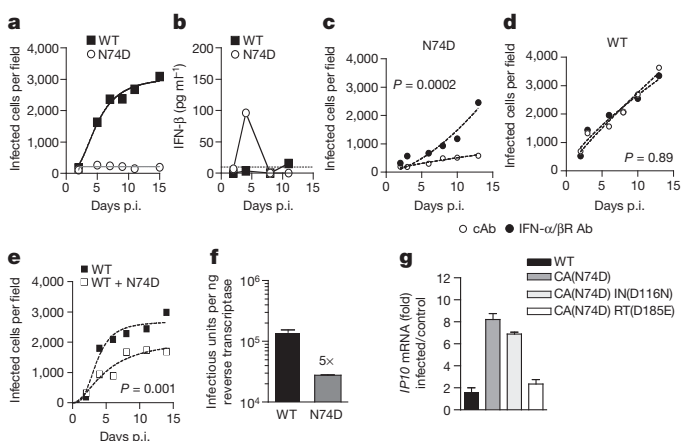


Figure 1 | HIV-1 CPSF6 binding mutant CA N74D is restricted in MDM due to induction of type 1 IFN. **a**, Replication of WT HIV-1 or CA mutant N74D in MDM. **b**, IFN- β levels in supernatants from **a**. **c**, **d**, Replication of HIV-1 CA N74D or WT HIV-1 with IFNAR2 or control antibody (cAb). **e**, Replication of WT or WT plus CA N74D. Mean data and regression lines for biological replicates are shown in **c–e**. **P** values (two-way ANOVA) are given for IFNAR2 blockade (**c–d**) and co-infection with CA mutant N74D (**e**). **f**, Infection of MDM by HIV-1 measured at 48 h. **g**, *GAPDH*-normalized *IP10* mRNA levels expressed as fold change over untreated cells after infection with WT or HIV-1 mutants (mean of 3 technical replicates \pm s.e.m., **f**, **g**).

¹University College London, Medical Research Council Centre for Medical Molecular Virology, Division of Infection and Immunity, University College London, 90 Gower Street, London WC1E 6BT, UK.

²Protein and Nucleic Acid Chemistry Division, Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. ³Wolfson Institute for Biomedical Research, University College London, Gower Street, London WC1E 6BT, UK.

*These authors contributed equally to this work.

Given that CA mutation N74D prevents recruitment of CPSF6^{1,3} we proposed that CPSF6 depletion would induce WT HIV-1 to trigger IFN responses in MDM. In fact, CPSF6 depletion by shRNA expression in MDM (Fig. 2a, b and Extended Data Fig. 2) completely abrogated HIV-1 replication (Fig. 2c) due to a burst of IFN- β . MDM expressing a non-targeting shRNA did not produce IFN- β on HIV-1 infection (Fig. 2d). The restrictive role of IFN was confirmed by rescue of infectivity with IFNAR2 antibody (Fig. 2e). Neither the IFNAR2 nor isotype control antibody had any effect on HIV-1 replication in control shRNA expressing MDM (Fig. 2f). Importantly, shRNA expression itself did not induce IFN- β production (Fig. 2g). We conclude that the defect in WT HIV-1 replication after CPSF6 depletion in MDM was largely due to type 1 IFN production. In line with observations made with HIV-1 N74D, CPSF6 depletion also reduced single-round infectivity in MDM by a few fold, 3.5-fold versus 5-fold (Fig. 2h).

The HIV-1 inhibitor PF-3450074 (PF74) binds CA and inhibits CPSF6 recruitment and HIV-1 replication^{1,3,8}. As expected, PF74 completely blocked HIV-1 replication in MDM but did not induce soluble IFN- β secretion, nor was replication rescued by IFNAR2 blockade (Extended Data Fig. 5a–c). However, as reported, PF74 completely abrogated HIV-1 DNA synthesis (Extended Data Fig. 5d, e)^{8,9}. The fact that PF74 mimics CPSF6 binding to HIV-1 CA³ suggests that CPSF6 recruitment might prevent premature reverse transcription and innate recognition of viral DNA. To test this hypothesis, a human CPSF6 mutant deleted for its nuclear localization signal (CPSF6 Δ NLS)¹⁰, was expressed in HeLa cells. Like PF74, human CPSF6 Δ NLS blocked VSV-G HIV-1 GFP DNA synthesis and infectivity (Extended Data Fig. 5f, g). A CPSF6 mediated block to HIV-1 RT differs from previous observations showing no effect of CPSF6 Δ NLS on HIV-1 RT but earlier work

used a mouse CPSF6 cDNA with an alternate exon structure^{1,4}. We hypothesize that HIV-1 has evolved to recruit CPSF6 to incoming HIV-1 CA and prevent premature DNA synthesis, which would otherwise trigger innate sensors (Extended Data Fig. 1).

Because HIV-1 N74D is unable to appropriately use nuclear pore components and has retargeted integration properties^{1–3,11}, we proposed that the HIV-1 CA mutant P90A, which fails to interact with the cyclophilins CypA and nuclear pore component Nup358, and also has retargeted integration², might also trigger innate sensors. Indeed, HIV-1 P90A infection of MDM induced IFN- β production and an antiviral state in both replication and single-round infectivity assays, which was rescued by IFN receptor antibody (Fig. 3a–f and Extended Data Figs 2 and 3c, d). We find MDM infection by HIV-1 N74D, P90A or WT were equally increased by macaque simian immunodeficiency virus-like particles (SIVmac VLP) encoding Vpx, indicating that mutant viruses were not specifically Vpx-sensitive (Extended Data Fig. 5h). Quantitative PCR with reverse transcription (qRT-PCR) and whole-genome profiling demonstrated ISG induction after HIV-1 P90A infection (Extended Data Fig. 4b–d). Consistently, double mutation of P90A and RT D185E, but not IN D116N, suppressed IP10 induction (Fig. 3g). We propose that viral DNA produced by reverse transcription is the target for innate sensing of both HIV-1 CA mutants N74D and P90A in MDM.

We next considered the mechanism of HIV-1 mutant innate sensing. A recently identified cytosolic DNA sensor, cyclic GMP-AMP synthase (cGAS), which synthesizes the novel second messenger cGAMP¹², has been shown to detect HIV-1 reverse-transcribed DNA in human myeloid cells¹³. cGAMP production is detected by stimulator of interferon genes (STING) that transduce an innate signalling cascade, leading to IRF3 activation and type 1 interferon production¹³. We used a biological assay to test for cGAMP production in MDM infected by HIV-1 N74D and P90A. Consistently, extracts from cells infected with HIV-1 CA(P90A) mutant contained a benzonase and heat-resistant component that activated an interferon-sensitive promoter in a STING-dependent way (Extended Data Fig. 6). Commercially prepared cGAMP validated the assay and acted as a positive control. Importantly, RNA purified from cells infected with HIV-1 mutant was not immunostimulatory, as oppose to RNA from cells infected with Sendai virus, which potentially activated an IFN- β promoter, as expected for a RIG-I-triggering virus.

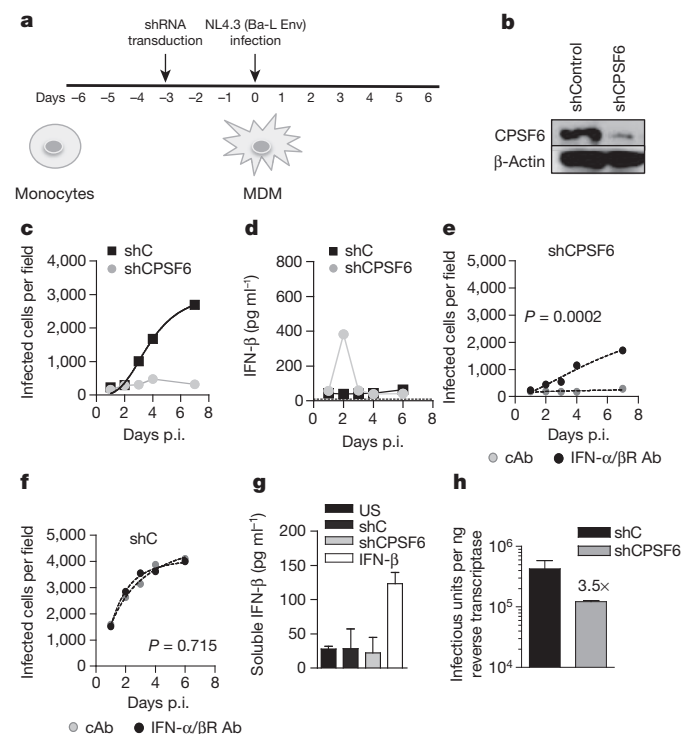


Figure 2 | HIV-1 elicits a type 1 IFN response that restricts replication in CPSF6 depleted MDM. **a**, Protocol schema. **b**, CPSF6/actin detected at time of infection. **c**, HIV-1 replication in MDM expressing shRNA targeting CPSF6 or control shRNA. **d**, IFN- β levels in supernatants from **c**. **e**, **f**, Infection of CPSF6-depleted or MDM expressing control shRNA with IFNAR2 or control antibody (cAb). P values (two-way ANOVA) are given for the effect of CPSF6 depletion (**e**) or control shRNA (**f**) on biological replicates. **g**, IFN- β produced from shRNA-expressing MDM or IFN- β -treated MDM. **h**, Infection of MDM by HIV-1 measured at 48 h on CPSF6-depleted or control shRNA-expressing MDM (mean of 3 technical replicates \pm s.e.m.).

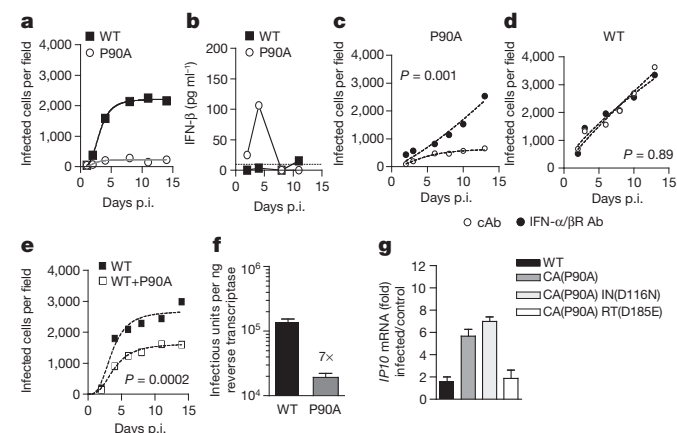


Figure 3 | HIV-1 CypA-binding mutant CA(P90A) is restricted in MDM owing to induction of type 1 IFN. **a**, Replication of WT HIV-1 or CA mutant P90A in MDM. **b**, IFN- β levels in supernatants from **a**. **c**, Replication of HIV-1 CA(P90A) with IFNAR2 or control antibody (cAb). **d**, As in Fig. 1d. **e**, Replication of WT or WT plus CA(P90A). Mean data and regression lines are shown for biological replicates in **c**–**e**. P values (two-way ANOVA) are given for IFNAR2 blockade (**c**, **d**) and co-infection with CA mutant P90A (**e**). **f**, Infection of MDM by HIV-1 measured at 48 h. **g**, GAPDH-normalized IP10 mRNA levels expressed as fold change over untreated cells after infection with WT or HIV-1 mutants (mean of 3 technical replicates \pm s.e.m.; **f**, **g**).

These data support our hypothesis that HIV-1 DNA is the pathogen-associated molecular pattern. Intriguingly, HIV-1 N74D did not stimulate in either of these assays, indicating that the two HIV-1 mutants activate independent DNA sensors. This possibility is consistent with the different integration site targeting preferences of the two mutants with HIV-1 N74D and P90A integrating into lower gene density or higher gene density regions of chromatin, respectively, compared to wild-type virus².

Immunofluorescent detection of NF- κ B and IRF3 revealed nuclear translocation of both transcription factors after exposure to either of the HIV-1 mutants but not WT virus (Fig. 4a–c and Extended Data Fig. 7). Concordantly, inhibition of NF- κ B activation with a peptide inhibitor of NEMO (IKK) rescued infectivity of both HIV-1 mutants in a dose-dependent manner (Fig. 4c). Finally, we considered whether prevention of cofactor interaction using drugs could induce WT virus to trigger a cell-autonomous innate immune response in the same way as mutant virus. We sought to phenocopy the HIV-1 P90A mutant by inhibiting cyclophilin recruitment using cyclosporine or a non-immunosuppressive analogue of cyclosporine, SmBz-CsA. SmBz-CsA is modified at the 3'-SAR position to include a methylphenyl-4-carboxylic

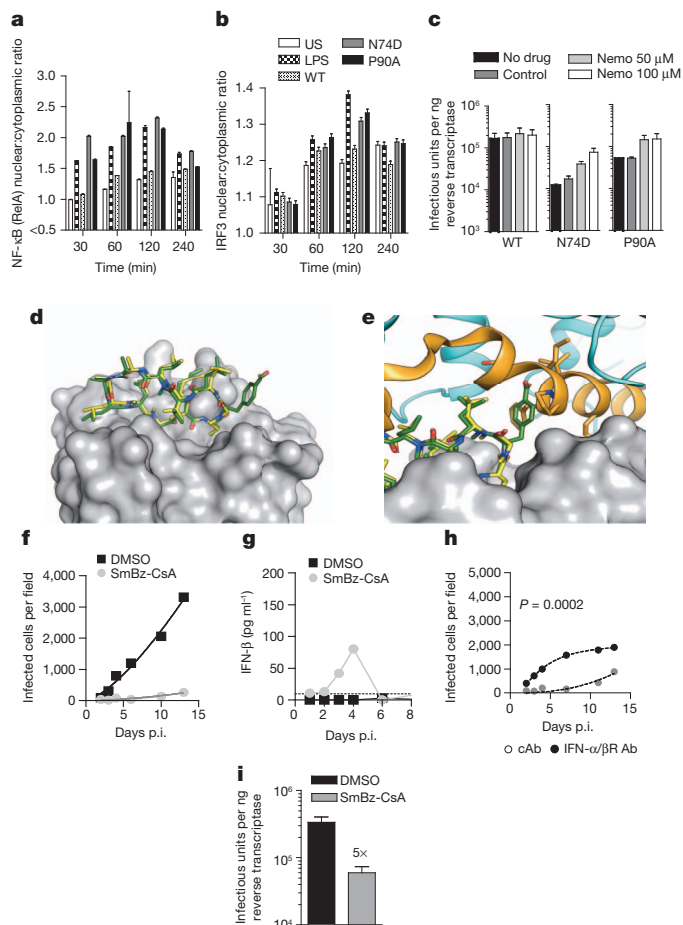


Figure 4 | NF- κ B/IRF3 are activated by mutant HIV-1 and SmBz-CsA treatment causes WT HIV-1 to trigger innate responses. **a, b,** Mean (\pm s.e.m.) nuclear/cytoplasmic ratios for NF- κ B or IRF3 in infected MDM ($P < 0.05$, two-way ANOVA)., LPS, lipopolysaccharide; Nemo, inhibitor peptide to IKK- α/β ; US, unstimulated. **c,** Infection at 48 h \pm IKK inhibitor. **d, e,** SmBz-CsA (green) complexed with CypA (grey), cyclosporine (yellow) and calcineurin (orange/blue). **f,** Replication of HIV-1 in MDM \pm SmBz-CsA; **g,** IFN- β levels from **f**. **h,** MDM infected with WT HIV-1 plus SmBz-CsA and IFNAR2 antibody or cAb (mean data and regression lines). P value (two-way ANOVA) is given for IFNAR2 blockade of biological replicates. **i,** Infection of MDM by WT HIV-1 \pm SmBz-CsA at 48 h (mean of 3 technical replicates \pm s.e.m.).

acid group and therefore cannot inhibit calcineurin or affect T cell activation¹⁴ (Fig. 4d, e and Extended Data Table 1). Like cyclosporine, SmBz-CsA inhibited recruitment of CypA, but not Nup358 Cyp, to HIV-1 CA (Extended Data Fig. 8e). Treatment of MDM with SmBz-CsA (Fig. 4f) or cyclosporine (Extended Data Fig. 8a) completely suppressed WT HIV-1 replication and elicited IFN- β production (Fig. 4g and Extended Data Fig. 8b). Inhibited viral replication was rescued by IFNAR2 blockade, but not by control antibody (Fig. 4h and Extended Data Fig. 8c). After single-round infection in the presence of either drug, infection was 5–7-fold lower (Fig. 4i and Extended Data Fig. 8d). These data are consistent with observations of cyclosporine inhibition of hepatitis C virus, in which innate immune responses are implicated¹⁵.

Our findings demonstrate that human macrophages are able to detect HIV-1 infection and activate a cell-autonomous innate immune signal, when specific interactions with HIV-1 cofactors are prevented by virus mutation (Figs 1 and 3), depletion of cofactor expression (Fig. 2) or pharmacological inhibition of cofactor recruitment (Fig. 4). We envisage that appropriate interaction between CA and CPSF6/cyclophilins normally allows evasion of innate sensors and promotes HIV-1 infection. We propose a model in which CPSF6/CypA recruitment to CA suppresses premature viral DNA synthesis and thus innate triggering. Inhibition of DNA synthesis by CPSF6 Δ NLS or the CPSF6 mimic PF74 support this possibility. In our model, nuclear entry of CPSF6 could release the virus, therefore enabling reverse transcription at the nuclear pore. The cytosolic exonuclease TREX1 degrades excess cytoplasmic DNA and prevents cGAS activation and IFN stimulation^{13,16} (Extended Data Fig. 8f–j). Our data indicate that DNA synthesized by HIV-1 mutants is insensitive to TREX1 degradation, either through nature or location, and in the case of HIV-1 CA(P90A), is detected by cGAS leading to cGAMP production. In MDDC, CypA has been suggested to have a different role, acting to aid detection of HIV-1 by innate sensors during egress¹⁷. These observations suggest that HIV-1 may rely on cell-type-specific cofactor use to protect it from innate immune defences. Intriguingly, both N74D and P90A CA mutants replicated in indicator cell lines GHOST and HeLa TZM-bl to WT levels (Extended Data Fig. 9). Replication was unaffected by IFN-receptor blockade and ISG expression was not induced, illustrating that these cell lines cannot respond in the same way to HIV-1 infection. They also suggest that the only obstacle to HIV-1 CA mutant replication in MDM is due to induction of innate responses. Our observations facilitate the further study of the relationship between HIV-1 and innate immunity. We envisage therapeutics, or vaccine adjuvants, which induce virus to trigger potent cell-autonomous innate immunity, IFN secretion and enhanced adaptive immune responses.

METHODS SUMMARY

Monocytes were isolated and incubated with macrophage colony stimulating factor (M-CSF) to induce macrophage differentiation. Full-length HIV-1 from molecular clones and virus-like particles (VLPs) were produced by transient transfection of HEK293T cells and purified by centrifugation through a sucrose cushion. Infections were performed by incubating 10^5 MDM per well in 48-well plates. Intracellular staining of p24 was performed at various time points post infection with anti-p24 antibodies followed by a LacZ-conjugated antibody and automated colony counting using an ELISPOT reader (AID). Type 1 interferon was measured by enzyme linked immunosorbent assay (PBL Interferon Source) in supernatants taken from the wells used to measure HIV-1 replication. Single-round HIV-1 infections of MDM were performed by measuring HIV-1 p24-positive cells, as above, 48 h post infection. In CPSF6 and TREX1 depletion experiments day 3 differentiating MDM were transduced with shRNA encoding HIV-1 vector and SIVmac VLPs encoding Vpx and day 6 cells were challenged with replication competent HIV-1. The cGAMP assay was performed by treating L929 cells stably expressing luciferase under the control of an interferon-sensitive response element with heat- and benzonase-treated extracts from infected MDM. Immunostimulatory RNA was assayed by sequentially transfecting 293T cells with a luciferase reporter under the control of the interferon beta promoter and immunostimulatory RNA purified from MDM infected with HIV-1, or Sendai virus as a positive control. Nuclear

translocation of IRF3 and NF- κ B was assessed by staining HIV-1 infected MDM with specific antibodies and measuring nuclear to cytoplasmic staining ratios.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 8 October 2013.

Published online 6 November; corrected online 20 November 2013 (see full-text HTML version for details).

- Lee, K. *et al.* Flexible use of nuclear import pathways by HIV-1. *Cell Host Microbe* **7**, 221–233 (2010).
- Schaller, T. *et al.* HIV-1 capsid-cyclophilin interactions determine nuclear import pathway, integration targeting and replication efficiency. *PLoS Pathog.* **7**, e1002439 (2011).
- Price, A. J. *et al.* CPSF6 defines a conserved capsid interface that modulates HIV-1 replication. *PLoS Pathog.* **8**, e1002896 (2012).
- Ambrose, Z. *et al.* Human immunodeficiency virus type 1 capsid mutation N74D alters cyclophilin A dependence and impairs macrophage infection. *J. Virol.* **86**, 4708–4714 (2012).
- Tsang, J. *et al.* HIV-1 infection of macrophages is dependent on evasion of innate immune cellular activation. *AIDS* **23**, 2255–2263 (2009).
- Besnier, C., Takeuchi, Y. & Towers, G. Restriction of lentivirus in monkeys. *Proc. Natl Acad. Sci. USA* **99**, 11920–11925 (2002).
- Iyer, S. R., Yu, D., Biancotto, A., Margolis, L. B. & Wu, Y. Measurement of human immunodeficiency virus type 1 preintegration transcription by using Rev-dependent Rev-CEM cells reveals a sizable transcribing DNA population comparable to that from proviral templates. *J. Virol.* **83**, 8662–8673 (2009).
- Shi, J., Zhou, J., Shah, V. B., Aiken, C. & Whitby, K. Small-molecule inhibition of human immunodeficiency virus type 1 infection by virus capsid destabilization. *J. Virol.* **85**, 542–549 (2011).
- Blair, W. S. *et al.* HIV capsid is a tractable target for small molecule therapeutic intervention. *PLoS Pathog.* **6**, e1001220 (2010).
- Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W. & Barabino, S. M. Distinct sequence motifs within the 68-kDa subunit of cleavage factor I_m mediate RNA binding, protein-protein interactions, and subcellular localization. *J. Biol. Chem.* **279**, 35788–35797 (2004).
- Ocwieja, K. E. *et al.* HIV integration targeting: a pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog.* **7**, e1001313 (2011).
- Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786–791 (2013).
- Gao, D. *et al.* Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341**, 903–906 (2013).
- Dube, H. *et al.* A mitochondrial-targeted cyclosporin A with high binding affinity for cyclophilin D yields improved cytoprotection of cardiomyocytes. *Biochem. J.* **441**, 901–907 (2012).
- Liu, J.-P., Ye, L., Wang, X., Li, J.-L. & Ho, W.-Z. Cyclosporin A inhibits hepatitis C virus replication and restores interferon-alpha expression in hepatocytes. *Transpl. Inf. Dis.* **13**, 24–32 (2011).
- Yan, N., Regalado-Magdos, A. D., Stiggelbout, B., Lee-Kirsch, M. A. & Lieberman, J. The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. *Nature Immunol.* **11**, 1005–1013 (2010).
- Manel, N. *et al.* A cryptic sensor for HIV-1 activates antiviral innate immunity in dendritic cells. *Nature* **467**, 214–217 (2010).

Acknowledgements We are grateful to J. W. Chin, S. Goodbourn, K. Lee, O. Perisic and V. KewalRamani for reagents and advice. This work was funded by Wellcome Trust Senior Fellowship 090940 to G.J.T., the Medical Research Council, an MRC Confidence in Concept Award to G.J.T. and D.S. and the National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Author Contributions J.R., C.P.T., A.J.F., D.L.S., M.N. and G.J.T. designed the study. J.R. performed all experiments in MDM, C.P.T. performed cGAMP and immunostimulatory RNA assays, A.J.F. performed CPSF6 experiments in HeLa cells. A.J.P. solved the structure of SmBz-CsC in complex with CypA and L.H. performed the TRIMCyp experiments. J.R., A.J.F., A.J.P., L.H., D.A.J., L.C.J., M.N. and G.J.T. analysed the data. A.J.F. and C.B. generated constructs and D.L.S. synthesized SmBz-Csa. J.R., M.N. and G.J.T. wrote the manuscript.

Author Information Structural coordinates have been deposited under PDB accession code 4IPZ. Microarray data are available from the EBI Array Express repository under accession no. E-MTAB-1437. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.T. (g.towers@ucl.ac.uk) and M.N. (m.noursadeghi@ucl.ac.uk).

METHODS

Cells. Primary monocyte-derived macrophages (MDM) were prepared from fresh blood from healthy volunteers as described⁵. The study was approved by the joint University College London/University College London Hospitals NHS Trust Human Research Ethics Committee and written informed consent was obtained from all participants. Briefly, peripheral blood mononuclear cells were isolated by Ficoll-Hypaque (Axis-Shield) density centrifugation. The isolated cells were washed with PBS and plated in RPMI (Invitrogen) supplemented with 10% heat-inactivated autologous human serum (HS) and 40 ng ml⁻¹ macrophage colony stimulating factor (M-CSF) (R&D systems). The medium was then refreshed after 3 days (RPMI 1640 with 10% HS), removing any remaining non-adherent cells. After 6 days, media was replenished with RPMI containing 5% type AB HS (Sigma-Aldrich). Replicate experiments were performed with cells derived from different donors.

GHOST¹⁸, a human osteosarcoma cell line stably expressing *CD4*, *CCR5*, *CXCR4* and the green fluorescent protein (*GFP*) reporter gene under the control of the HIV-2 long terminal repeat were maintained in DMEM containing 10% heat-inactivated fetal calf serum (FCS), glutamine, antibiotics, G418 (500 µg ml⁻¹), hygromycin (100 µg ml⁻¹), and puromycin (1 µg ml⁻¹) and were split twice a week. HEK293T cells were grown in DMEM (Invitrogen) supplemented with 10% FCS.

Reagents. Recombinant human interferon (IFN)-β (Merck Serono) was used at 10 ng ml⁻¹, poly(I:C) (Sigma) was used at 10 µg ml⁻¹, Cyclosporine (Sandoz) was used at 5 µM. SmBz-CsA was synthesized as described¹⁴ and used at 10 µM. PF74, a gift from J. Chin, was synthesized as described³ and used at 10 µM. Lipopolysaccharide (LPS) (Sigma) was used at 100 ng ml⁻¹. Commercially prepared STING agonist, cyclic GMP-AMP (cGAMP), was purchased from InvivoGen.

Plasmids. The CCR5-tropic wild-type NL4.3 (Ba-L Env) or NL4.3 (Ba-L Env) bearing CA mutations 90A or N74D were derived from an infectious clone of NL4.3 by cloning the *Env* gene from HIV-1 Ba-L between unique EcoRI and BamHI sites to replace the NL4.3 *Env* gene. ΔRT and ΔIN infectious clones were generated by making mutant RT(D185E)⁶, or IN(D116N)⁷ using site-directed mutagenesis (Stratagene).

Short hairpin sequences were expressed from HIV-1-based shRNA expression vector HIVSiren². *CPSF6* shRNA target sequence was 5'-CGAAGAGTTCAACC AGGAA-3'; *TREX1* shRNA target sequence was 5'-CCAAGACCATTCTGCTG TCA-3'; *CPSF6* was detected by western blot and *TREX1* was detected by qRT-PCR.

A human *CPSF6* expression vector was prepared by PCR cloning the human *CPSF6* ORF from complementary DNA (Superscript, Life Technologies), prepared from HeLa cells, into the MLV based gammaretroviral expression vector EXN¹⁹ using primers forward 5'-ATCGGAATTCATGGCGGACGGTGTGGAC CACATAGACATTAC-3' and reverse 5'-ATGCGCGGCCGCTTAACGATG ACGATATTCGCGCTCTC-3', restriction sites underlined. The nuclear localization signal was removed from *CPSF6* as described¹⁰ by deleting the C-terminal 50 amino acids by PCR using reverse primer 5'-ATGCGCGGCCGCTCATTCT CGTGATCTACTATGGTCCC-3' and forward primer as above. The resulting NLS mutant is defective for nuclear entry as described¹⁰. *CPSF6*ΔNLS was expressed in HeLa cells by gammaretroviral vector transduction as described³ and G418-selected pools of cells generated. Note that the human *CPSF6* cDNA described herein differs from the murine cDNA described previously^{1,4,20} in that it represents the most common human *CPSF6* isoform represented by GenBank accession number nm007007 and thus lacks exon 6^{21–23}.

Virus production. Virus particles were produced by transient transfection of HEK293T cells. 3.5 µg of molecular clone DNA; for shRNA we used 1.5 µg pHIVSIREN² shRNA, 1 µg p8.91²⁴ and 1 µg pMDG²⁵ encoding VSV-G protein. For SIVmac-VLP we left out the genome plasmid and transfected 3 µg pSIV3+²⁶ and 1 µg pMDG using 10 µl FuGENE 6 transfection reagent (Promega) as described⁶. HIV-1 GFP was produced by transfection of 293T with GFP-encoding genome CSGW, packaging plasmid p8.91 and pMDG as described⁶. Virus supernatants were collected 48 h, 72 h and 96 h post transfection. All virus suspensions were filtered and ultracentrifuged through a 20% sucrose buffer and resuspended in RPMI 1640 with 5% HS, for subsequent infection of MDM. All virus preparations were quantified by reverse transcriptase (RT) enzyme-linked immunosorbent assay (ELISA) (Roche) except when doses were measured by p24 CA ELISA (National Cancer Institute at Frederick) where stated (Figs 1g and 3g). Viruses were also titrated on GHOST where described detecting infection by flow cytometry 72 h post infection or HeLa TZM bl where infection was detected by CA staining as below.

Infection and stimulation. MDM were infected with 100 pg reverse transcriptase enzyme-linked immunosorbent assay (RT-ELISA) (Roche) per well (multiplicity of infection (MOI) 0.2) in 48-well plates and subsequently fixed and stained using CA-specific antibodies (EVA365 and EVA366 National Institute of Biological Standards AIDS Reagents Programme) and a secondary antibody linked to beta

galactosidase, as described⁵. During the time course, supernatants were collected for IFN-β ELISA (PBL Interferon Source) according to manufacturer's instructions. Anti-IFN-α/β receptor (PBL Interferon Source) or control IgG2A antibody (R&D systems) were added at 1 µg ml⁻¹ for 2 h before infection and supplemented every 4 days. For inhibition of NF-κB activation, a peptide inhibitor of NEMO (IKKγ) or control peptide (Imgenex), were added at either 50 µM or 100 µM for 12 h before infection. For Agilent microarray analysis and qRT-PCR, MDM were infected with 1 ng RT per well (MOI 2) in 24-well plates. RNA was extracted 24 h post infection (RNeasy, Qiagen) and subject to microarray analysis as described later. For shRNA transduction of MDM, day 3 differentiated cells were infected with shRNA (0.1 ng RT per ml), SIVmac-VLP (1 ng RT per ml) + 8 µg ml⁻¹ polybrene overnight.

Western blot analysis. *CPSF6* expression was measured in extracted cell pellets by western blot. Cells were lysed in Laemmli buffer then boiled before separation by SDS-PAGE as described previously³. After *CPSF6* or STING detection membranes were stripped and probed again for β-actin as a loading control. Antibodies used were *CPSF6* (Abcam ab99347) STING (Abcam ab82960) and β-actin (Abcam ab6276).

Microarray analysis. Total RNA was purified from cell lysates collected in RLT buffer (Qiagen) using the RNeasy Mini kit (Qiagen). Samples were processed for Agilent microarrays as previously described⁵ and loess normalized data were analysed using the TM4 microarray software suite MeV v4.8²⁷. Pathway enrichment analysis of differentially expressed gene lists was performed using the online bioinformatics tool InnateDB²⁸. Microarray data are available from the EBI Array Express repository (<http://www.ebi.ac.uk/arrayexpress/>) under accession no E-MTAB-1437.

Quantitative PCR. cDNA was synthesized using the Omniscript RT Kit (Qiagen) and quantitative PCR of selected genes was performed using the following inventoried TaqMan assays (Applied Biosystems) *CCL8* (Hs04187715_m1) and *IFIT1* (Hs01911452_s1). *IP10* expression was quantified using: forward primer: 5'-TGA AATTATCTCTGCAAGCCAATT-3', reverse primer: 5'-CAGACATCTCTTCT CACCTTCTTT-3', and probe: 5'-TGTCCACGTGTTGAGATCATTCATGCTACA ATG-3'. *TREX1* expression was quantified using: forward primer: 5'-GCATCTG TCAGTGAGACCA-3', reverse primer: 5'-AGATCCTTGGTACCCCTGCT-3', and probe: 5'-CACAAACCAGGAACACTAGTCCAGC-3'. Expression levels of target genes were normalized to glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) as previously described⁵. To measure late reverse transcription products, total DNA was purified 9 h post infection (QIAamp, Qiagen) with DNase-treated virus (70 U ml⁻¹ DNase (Affymetrix) in RQ1 buffer (Promega) for 37 °C, 1 h) and 500 ng were subjected to TaqMan quantitative PCR using late reverse transcription primers and probe to detect provirus as described²⁹. Cells were infected with virus that had been boiled for 2 min as a negative control. Infectivity was measured in parallel samples by intracellular p24 staining 48 h post infection. Presented qPCR experiments are means of technical replicates and represent 3 biological replicates.

cGAMP reporter assay. MDM were infected with 1 ng RT per well (MOI 2) in 24-well plates for 18 h. Cells were lysed in hypotonic buffer (10 mM Tris pH 7.4, 10 mM KCl and 1.5 mM MgCl₂). After freeze thaw, a proportion of cell extract was kept for stimulatory RNA reporter assay and the remaining was heated to 96 °C for 10 min. Sonicated extracts were centrifuged (20,000g, 20 min, 4 °C), followed by benzonase treatment (1 U µl⁻¹ benzonase (Novagen), 2 mM ATP, 37 °C 90 min). 4 µl of lysate was introduced to reporter cells using Lipofectamine 2000 (Invitrogen). The reporter cells are L929 cells or L929 cells depleted of STING by transfecting (Oligofectamine, Invitrogen) a previously described STING siRNA³⁰. The L929 cells stably express firefly luciferase driven by an interferon sensitive response element. Luciferase was read after 16 h using Steady-Glo (Promega) and a luminometer.

Immunostimulatory RNA reporter assay. Cell extracts were subjected to TRIzol extraction (Invitrogen) and the extracted RNA, plus a control plasmid encoding *Renilla* luciferase, was transfected into 293T cells expressing firefly luciferase driven by IFN-β promoter. Cells were transfected with 500 ng of RNA extracted from macrophages and luciferase values were determined at 16 h using dual-luciferase assay kit (Promega). IFN-β promoter activity (firefly luciferase) was normalized by global transcription (*Renilla* luciferase) and fold induction compares normalized luciferase values against mock-transfected reporter cells. As a positive control we infected MDM with Sendai virus, a gift from S. Goodbourn, and purified immunostimulatory RNA. All transfections in this assay use Lipofectamine 2000 (Invitrogen).

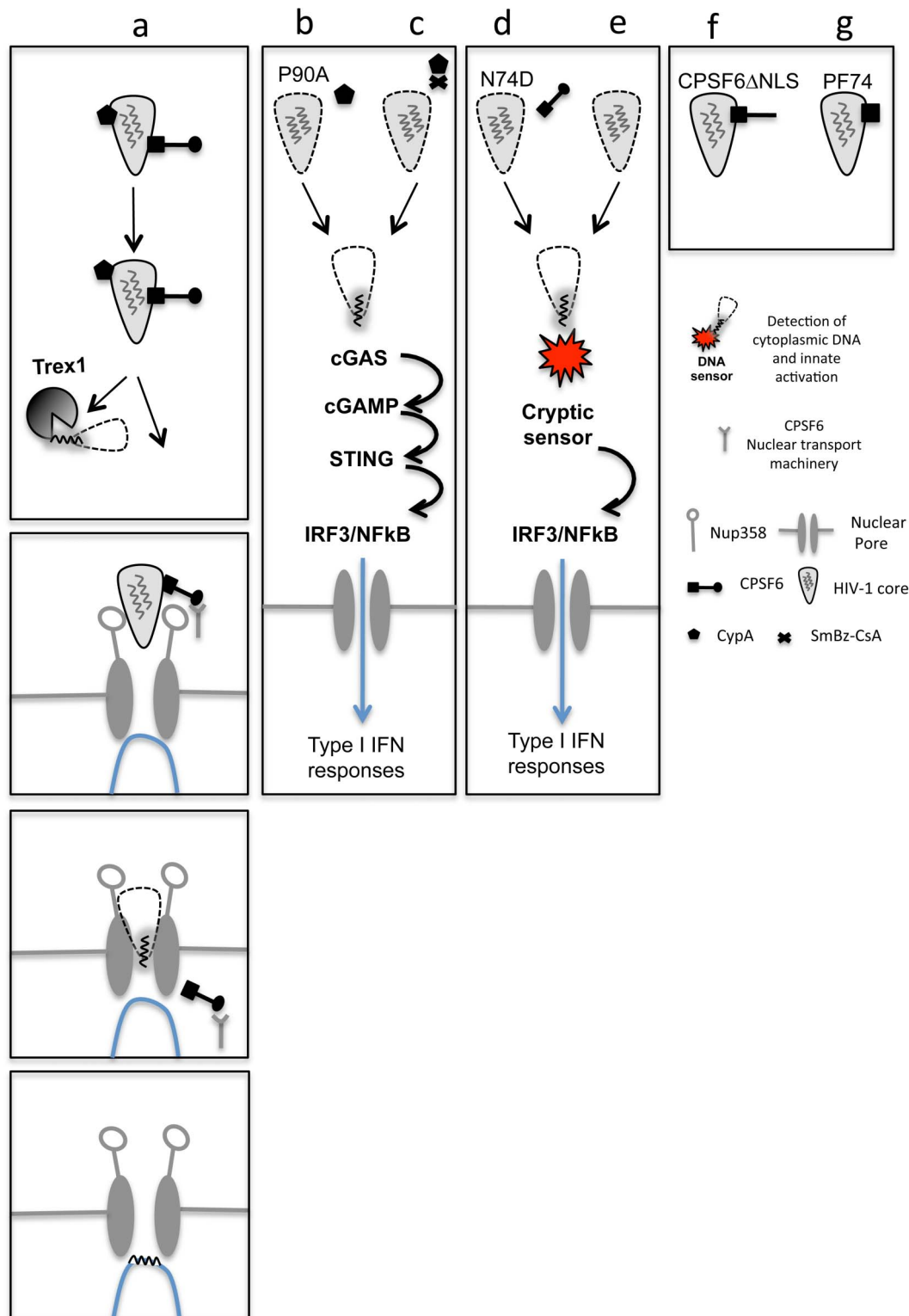
Quantitative confocal immunofluorescence analysis of NF-κB and IRF3 nuclear translocation. Nuclear/cytoplasmic ratios of NF-κB RelA and IRF3 transcription factors were analysed as previously described⁵ using a Hermes WiScan Cell Imaging System to analyse cells stained with rabbit polyclonal anti NF-κB

RelA (clone C-20) (Santa Cruz Biotechnology) or rabbit polyclonal anti IRF3 (clone FL 425) (Santa Cruz Biotechnology).

Protein expression, purification, crystallization, data collection, structure determination and refinement. CypA was expressed in *Escherichia coli* C41(DE3) cells (Lucigen) from tagless expression vector pOPT³. Cells were grown overnight at 18 °C before being collected, sonicated and purified by SP ion-exchange chromatography (GE Healthcare) followed by gel filtration. Crystals of SmBz-CsA in complex with CypA were grown at 17 °C in sitting drops. Protein solution (1 mM each of CypA and SmBz-CsA in 20 mM Tris pH 8, 50 mM NaCl, 1 mM DTT, 1% DMSO) was mixed with reservoir solution (1 M LiCl, 0.1 M MES pH 6, 30% w/v PEG 6000) in a 1:1 mix, producing 0.15 mm × 0.10 mm × 0.10 mm crystals within 24 h. Crystals were flash-frozen in liquid nitrogen before data collection using an in-house Mar-345 detector. Crystal data and diffraction statistics are provided in Extended Data Table 1. Crystallographic analysis was performed using programs from the CCP4 suite³¹. Data were indexed and scaled in MOSFLM and SCALA, respectively. The structure of SmBz-CsA:CypA (pdb 1CWA³²) was used as a search model. Structures were refined in REFMAC and Coot³¹. Structural figures were created using PyMol (<http://pymol.sourceforge.net/>). PDB coordinates have been deposited under accession code 4IPZ.

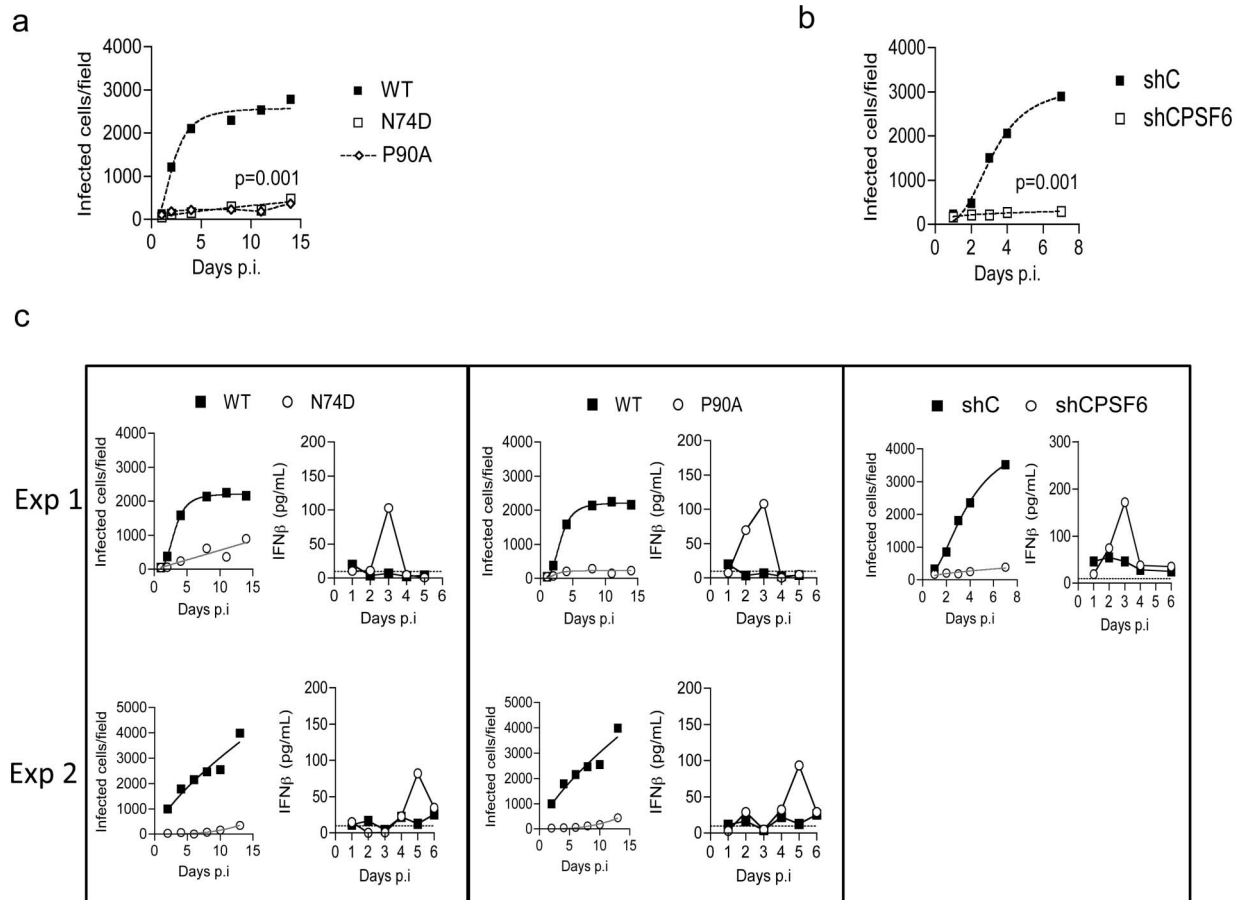
Statistical methods. All data were normally distributed and analysed for statistically significant differences between experimental groups by *t*-tests or two-way ANOVA as indicated. Bar charts show mean ± s.e.m. for experimental replicates in each case. Replication assays are presented for individual experiments, or where *P* values (two-way ANOVA) are given, replicate experiments. Individual or mean data points and nonlinear regression lines are shown over time. Sample sizes for each experiment were based on pilot experiments to estimate the effect size and variance of the data.

18. Mörner, A. *et al.* Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. *J. Virol.* **73**, 2343–2349 (1999).
19. Zhang, F., Hatzioannou, T., Perez-Caballero, D., Derse, D. & Bieniasz, P. D. Antiretroviral potential of human tripartite motif-5 and related proteins. *Virology* **353**, 396–409 (2006).
20. Lee, K. *et al.* HIV-1 capsid-targeting domain of cleavage and polyadenylation specificity factor 6. *J. Virol.* **86**, 3851–3860 (2012).
21. Ruepp, M. D., Schumperli, D. & Barabino, S. M. mRNA 3' end processing and more—multiple functions of mammalian cleavage factor I-68. *Wiley interdisciplinary reviews. RNA* **2**, 79–91 (2011).
22. Yang, Q., Gilmartin, G. M. & Doublet, S. The structure of human cleavage factor I_m hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5' capping and splicing. *RNA Biol.* **8**, 748–753 (2011).
23. Hori, T. *et al.* A carboxy-terminally truncated human CPSF6 lacking residues encoded by exon 6 inhibits HIV-1 cDNA synthesis and promotes capsid disassembly. *J. Virol.* **87**, 7726–7736 (2013).
24. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery *in vivo*. *Nature Biotechnol.* **15**, 871–875 (1997).
25. Naldini, L. *et al.* In vivo gene delivery and stable transduction of non-dividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).
26. Nègre, D. *et al.* Characterization of novel safe lentiviral vectors derived from simian immunodeficiency virus (SIVmac251) that efficiently transduce mature human dendritic cells. *Gene Ther.* **7**, 1613–1623 (2000).
27. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).
28. Lynn, D. J. *et al.* Curating the innate immunity interactome. *BMC Syst. Biol.* **4**, 117 (2010).
29. Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration *in vivo*. *Nature Med.* **7**, 631–634 (2001).
30. Ishikawa, H., Ma, Z. & Barber, G. N. STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* **461**, 788–792 (2009).
31. Collaborative Computational Project, number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
32. Mikol, V., Kallen, J., Pflugl, G. & Walkinshaw, M. D. X-ray structure of a monomeric cyclophilin A-cyclosporin A crystal complex at 2.1 Å resolution. *J. Mol. Biol.* **234**, 1119–1130 (1993).



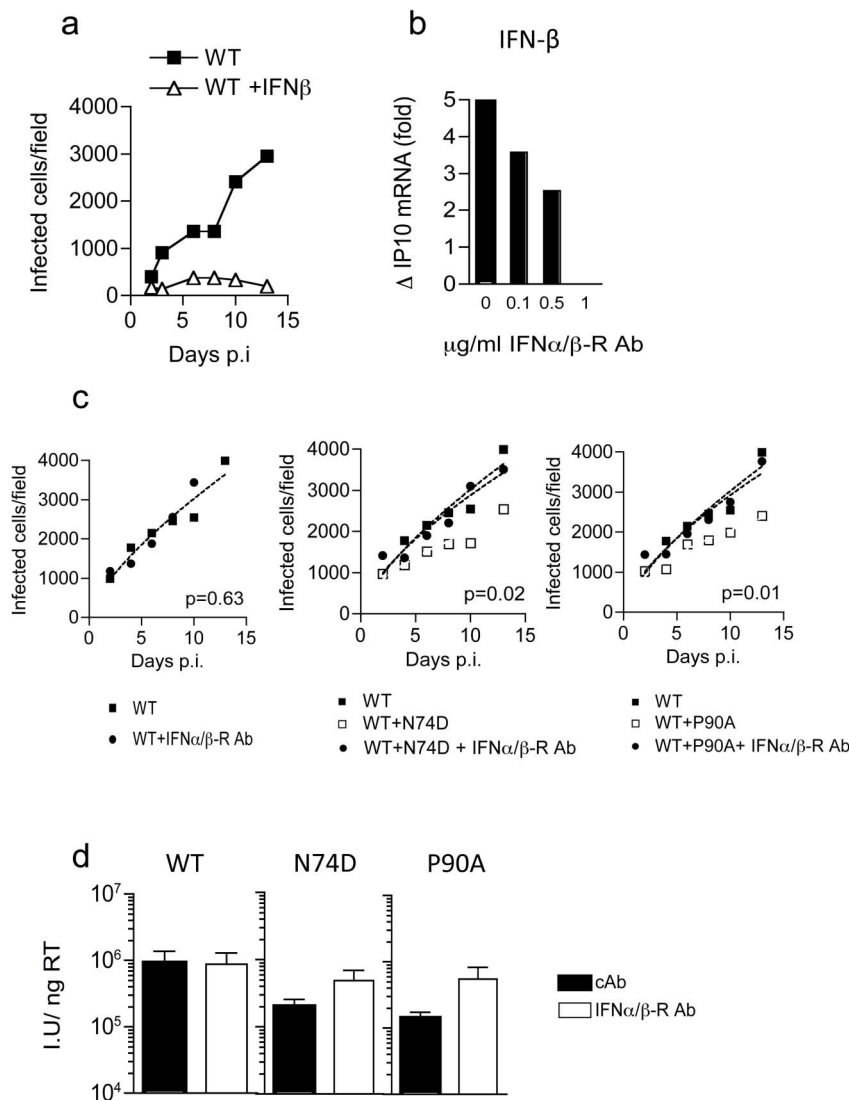
Extended Data Figure 1 | A model for HIV-1 core behaviour and innate sensing. **a**, Intact capsids recruit CypA and CPSF6 which direct the virus to the nucleus. CPSF6 interaction prevents premature DNA synthesis. Excess cytoplasmic DNA is degraded by TREX1. At the nuclear pore CPSF6 NLS-dependent dissociation from the virus allows reverse transcription to proceed. Reverse-transcribed DNA crosses the nuclear membrane and integrates. **b**, **c**, Disruption of CypA–CA interactions with either CA(P90A) mutation or cyclosporine treatment leads to detection of DNA reverse transcription product by cGAS initiating cGAMP production, STING activation, NF-κB/IRF3 nuclear localization, type I interferon secretion and initiation of an

antiviral state. **d**, **e**, Disruption of CPSF6–CA interactions by N74D CA mutation, or depletion of CPSF6, leads to activation of a cryptic innate DNA sensor which also activates NF-κB/IRF3 nuclear localization. **f**, Disruption of CPSF6 engagement with the nuclear transport machinery by mutating its NLS prevents reverse transcription because the CPSF6 does not dissociate from the capsid at the nuclear pore. **(g)** PF74 mimics CPSF6 by inserting a phenyl ring into a CA pocket in the same position as CPSF6 and also prevents reverse transcription. Like CPSF6ΔNLS PF74 has no NLS and thus does not disengage from the core and therefore terminally prevents reverse transcription.



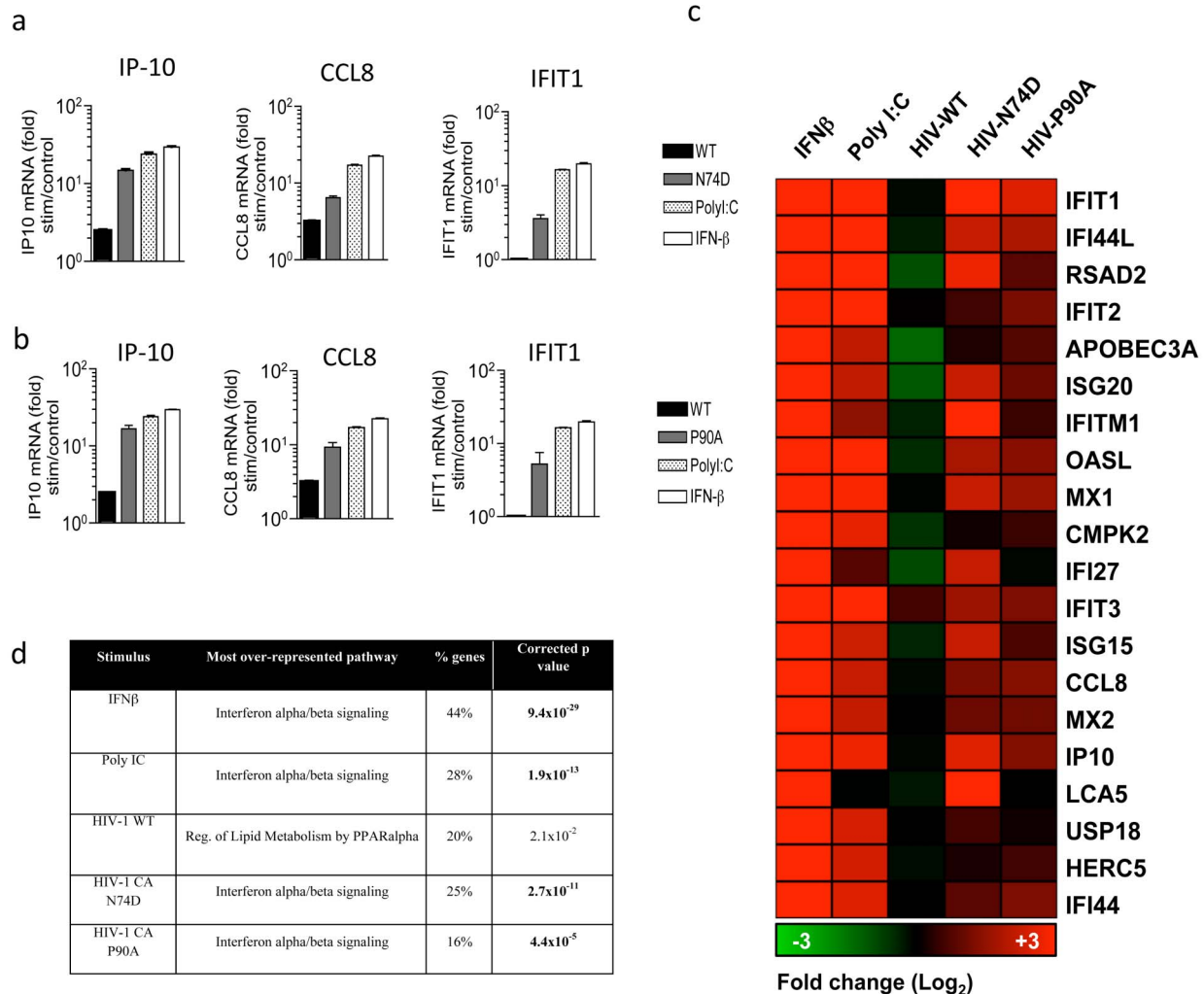
Extended Data Figure 2 | HIV-1 mutants CA N74D and P90A, or WT HIV-1 on CPSF6 depletion, induce Type I IFN secretion in human macrophages that limits propagation. **a**, MDM were infected with HIV-1 WT, CA N74D or CA(P90A) at low multiplicity. Cells were stained for Gag p24 at specific time points after infection and infected colonies counted. **b**, MDM transduced to express shRNA targeting CPSF6, or a scrambled control hairpin, were infected with wild-type NL4.3 (Ba-L Env) at low multiplicity. Cells were stained for Gag

p24 at specific time points after infection and infected colonies counted. (a-b) $P = 0.001$, two-way ANOVA, for the effect of CA mutation or CPSF6 depletion. Data represent mean and nonlinear regression over time for 3 biological replicates. **c**, Individual experiments from two additional donors performed as experiments shown in Fig. 1a and b, and one additional donor performed as experiments in Fig. 2c and d. Each HIV-1 mutant is shown compared to wild-type virus data (WT) for comparison.



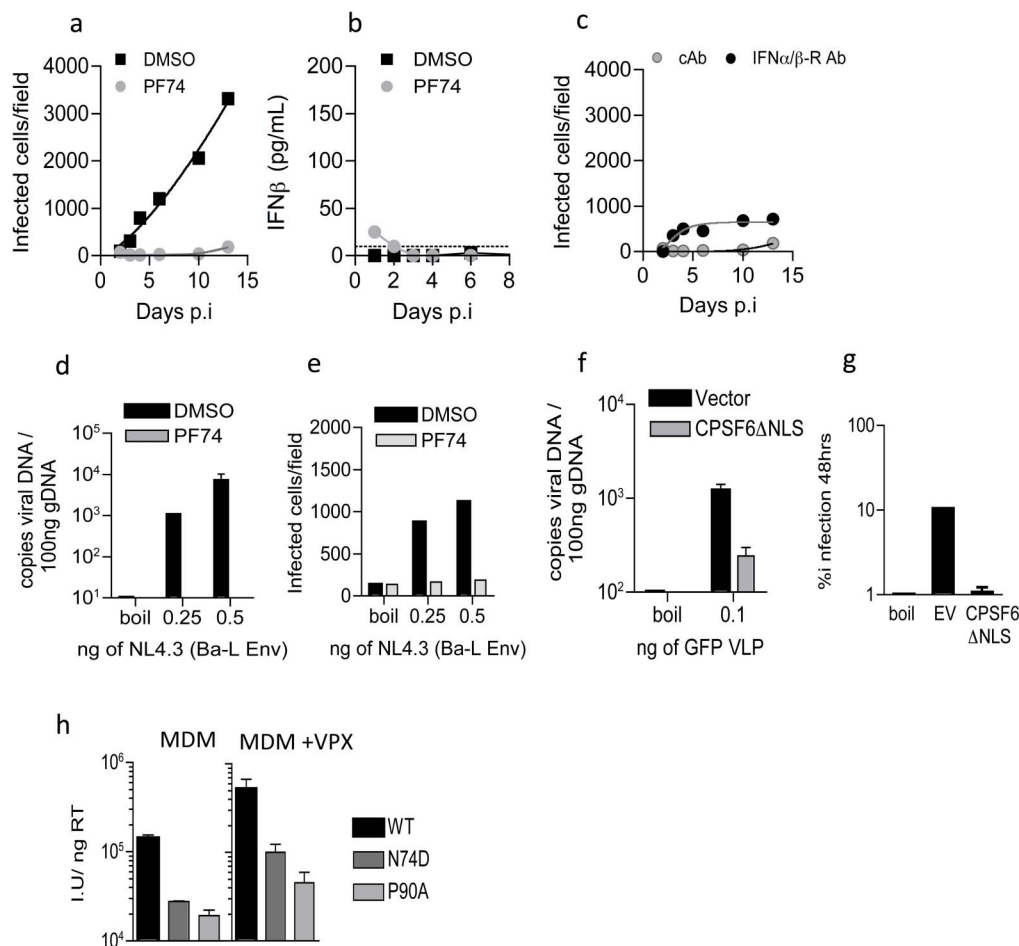
Extended Data Figure 3 | Suppression of HIV-1 by type 1 interferon and rescue of infectivity with anti-IFN receptor (IFNAR2) antibody. **a**, In order to demonstrate that IFN is able suppress wild type HIV-1 replication, MDM were pretreated with 1 ng ml^{-1} of recombinant IFN- β for 2 h then infected with HIV-1 WT NL4.3 (BaL-Env) infection. Cells were stained for Gag p24 at specific time points post infection (p.i). **b**, In order to determine how much IFN- α /β receptor (IFNAR2) neutralizing antibody is required to neutralize an IFN response, MDM were pretreated with varying concentrations of anti-IFNAR2 antibody for 2 h then stimulated with 1 ng ml^{-1} of recombinant IFN- β for 24 h. IP10 gene expression levels were measured by qRT-PCR and normalized to GAPDH. Results are expressed as fold change of expression over untreated cells. 1 μg ml^{-1} of IFNAR2 antibody effectively neutralized

1 ng ml^{-1} recombinant IFN- β , and this dose was used in subsequent experiments. **c**, MDM were infected with WT or WT and CA mutants at low multiplicity in the presence of anti-IFNAR2 antibody. Cells were stained for Gag at specific time points after infection and infected colonies counted (P values are given for two-way ANOVA for the effect of IFNAR blockade). Data represent mean and nonlinear regression of biological replicate experiments over time. **d**, Infectious titres of WT and CA mutant viruses were determined on MDM measured by assay of p24 positive cells 48 h post infection. Cells were infected in the presence of anti-IFNAR2 antibody or isotype control antibody (cAb). Titres are expressed as infectious units per nanogram of reverse transcriptase activity determined by ELISA. Mean \pm s.e.m. of titre determined at 3 doses (technical replicates).



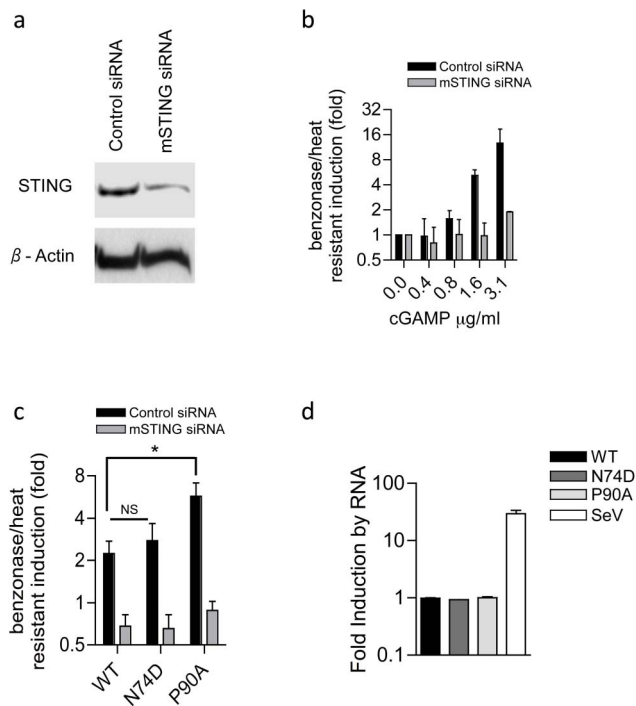
Extended Data Figure 4 | Stimulation of gene expression by HIV-1 CA mutants in MDM. **a, b**, Selected IFN-stimulated genes significantly upregulated by HIV-1 CA mutant infection, as well as by IFN- β and poly(I:C) measured at 24 h shown as fold change in expression (Stim/Control) measured by qRT-PCR and normalized to GAPDH mRNA levels. **c**, The same RNA samples as **a, b** were subjected to expression array and are presented in an expression matrix illustrating fold change in gene expression. **d**, Upregulation of gene expression (mean >2 fold in 2 independent biological replicates) after infection of MDM by HIV-1 wild-type and mutants HIV-1 CA N74D and

HIV-1 CA(P90A) as shown. 24 h after infection (MOI 2), total RNA was isolated and subject to expression array, see methods. Results were subject to pathway analysis using the online bioinformatics tool- InnateDB (<http://www.innatedb.com>). Type 1 IFN signalling was the most significantly over-represented pathway with IFN- β , PolyIC and both HIV-1 mutants, but not WT virus, based on the Reactome database (<http://www.reactome.org>). The proportion of genes in each list that map to this pathway and the p-value following Benjamini-Hochberg correction for multiple biological replicates are indicated.



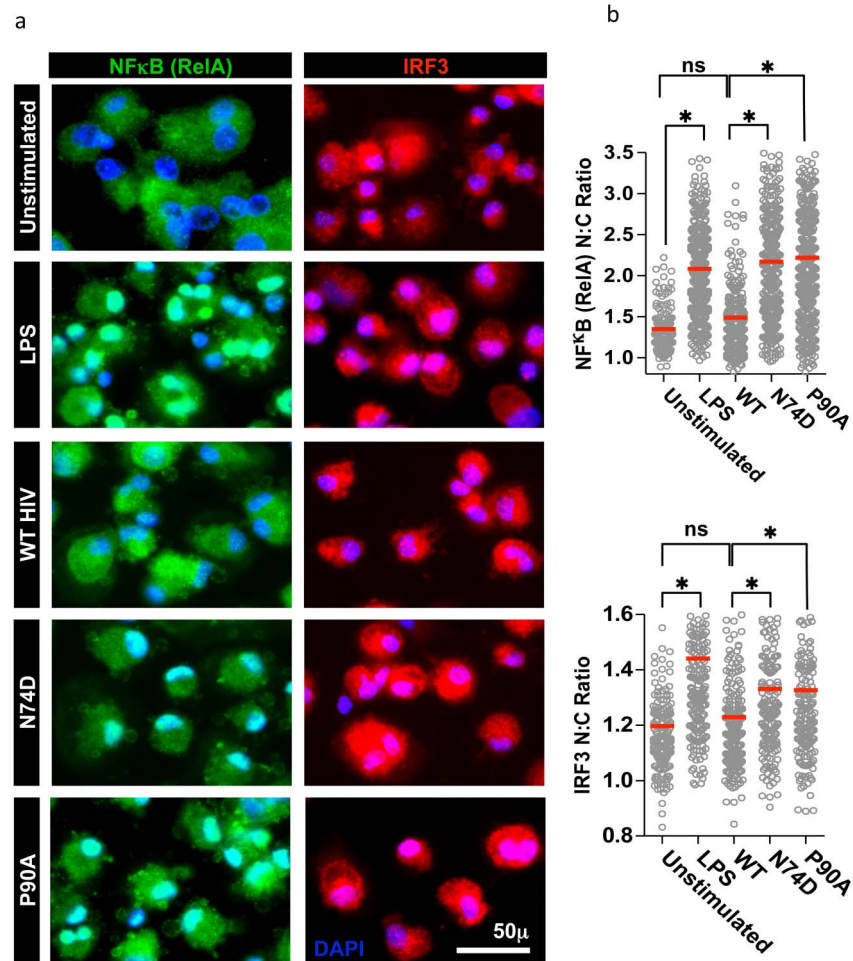
Extended Data Figure 5 | Inhibition of HIV-1 with PF74 in MDM does not trigger IFN production. **a**, Infection of human MDM with WT NL4.3 (Ba-L Env) in the presence or absence of 10 μ M PF74. Cells were stained for p24 at specific time points after infection and infected colonies counted. **b**, Supernatants collected from MDMs in **a** were assayed for soluble IFN- β levels by ELISA. **c**, MDM were infected at low multiplicity with WT HIV-1 in the presence of PF74 and IFNAR2 antibody or isotype cAb. **d**, Measurement of HIV-1 late reverse transcription product (LRT) in MDM infected for 9 h with two concentrations of WT HIV-1, in the presence or absence of 10 μ M PF74. Cells infected with boiled virus served as a negative control for DNA contamination. **e**, A sample parallel to those in **d** was used to determine the

number of infected cells by staining for p24 48 h post infection. **f**, Measurement of HIV-1 LRT product in HeLa cells that express CPSF6 Δ NLS or empty vector (EV), infected for 9 h with VSV-G pseudotyped HIV-1 GFP. **g**, A parallel sample was used to determine the number of infected cells by flow cytometry 48 h post infection. **h**, Infectious titres of WT and CA mutant viruses were determined on MDM in the presence and absence of SIVmac-VLPs encoding the SAMHD1 antagonist Vpx. Titres are expressed as infectious units per nanogram of reverse transcriptase activity determined by ELISA. Mean \pm s.e.m. of titre determined at 3 doses. Experiments represent 3 independent biological replicates.



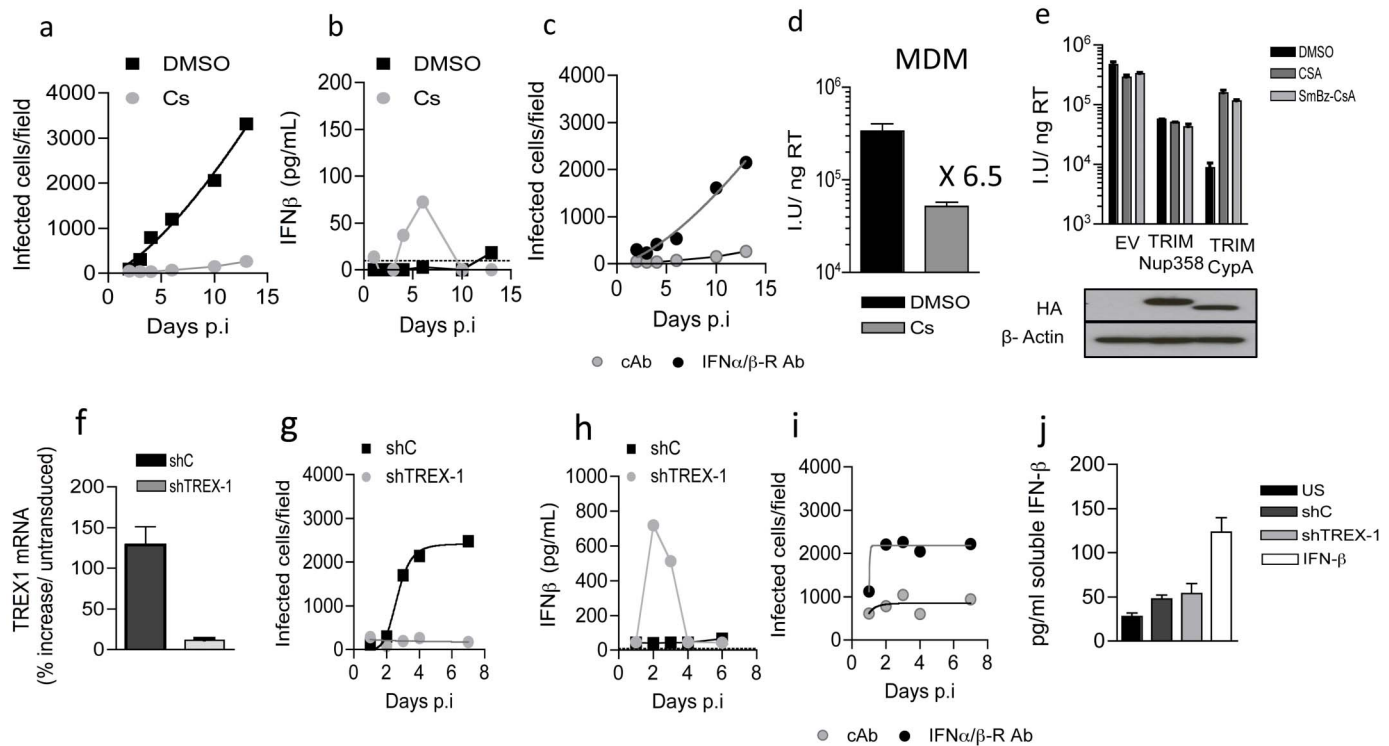
Extended Data Figure 6 | HIV-1 P90A CA mutant infected MDM contained a benzonase and heat resistant component that activated an interferon sensitive promoter.

a, Immunoblot of extracts from L929 cells stably expressing Firefly Luciferase-driven by an interferon sensitive response element depleted of STING by siRNA or expressing control siRNA. **b**, L929 cells, control or STING depleted as in **a**, were treated with various concentrations of STING agonist cyclic GMP-AMP (cGAMP), and Luciferase activity was read after 16 h. **c**, MDM were infected with HIV-1 wild-type (WT) or mutant HIV-1 CA N74D or HIV-1 CA(P90A) for 18 h (MOI of 2), total cell extracts were isolated and were heat and benzonase treated and were applied to L929 cells as in **a** and Luciferase activity measured at 16 h (Mean of $4 \pm$ s.e.m. of biological replicates). **d**, As **c** except RNA was extracted from the cell extracts and were transfected into 293T cells to measure IFN- β promoter driven Luciferase activity after 16 h. Sendai virus infection served as positive control for immunostimulatory RNA (Mean \pm s.e.m. of biological replicates). **c**, * represents statistically significant difference between data sets ($P < 0.05$, t -test), NS represents non-significant differences.



Extended Data Figure 7 | Nuclear translocation of NF-κB and IRF3 after HIV-1 capsid mutant infection in MDM. **a**, Confocal immunofluorescence microscopy was used to quantify nuclear translocation of NF-κB Rel A (green) and IRF3 (red) as a consequence of activation. Nuclear:cytoplasmic ratios of immunostaining were measured at single cell level by quantitation of NF-κB or IRF3 signal intensities inside and outside the nucleus (blue DAPI) as described

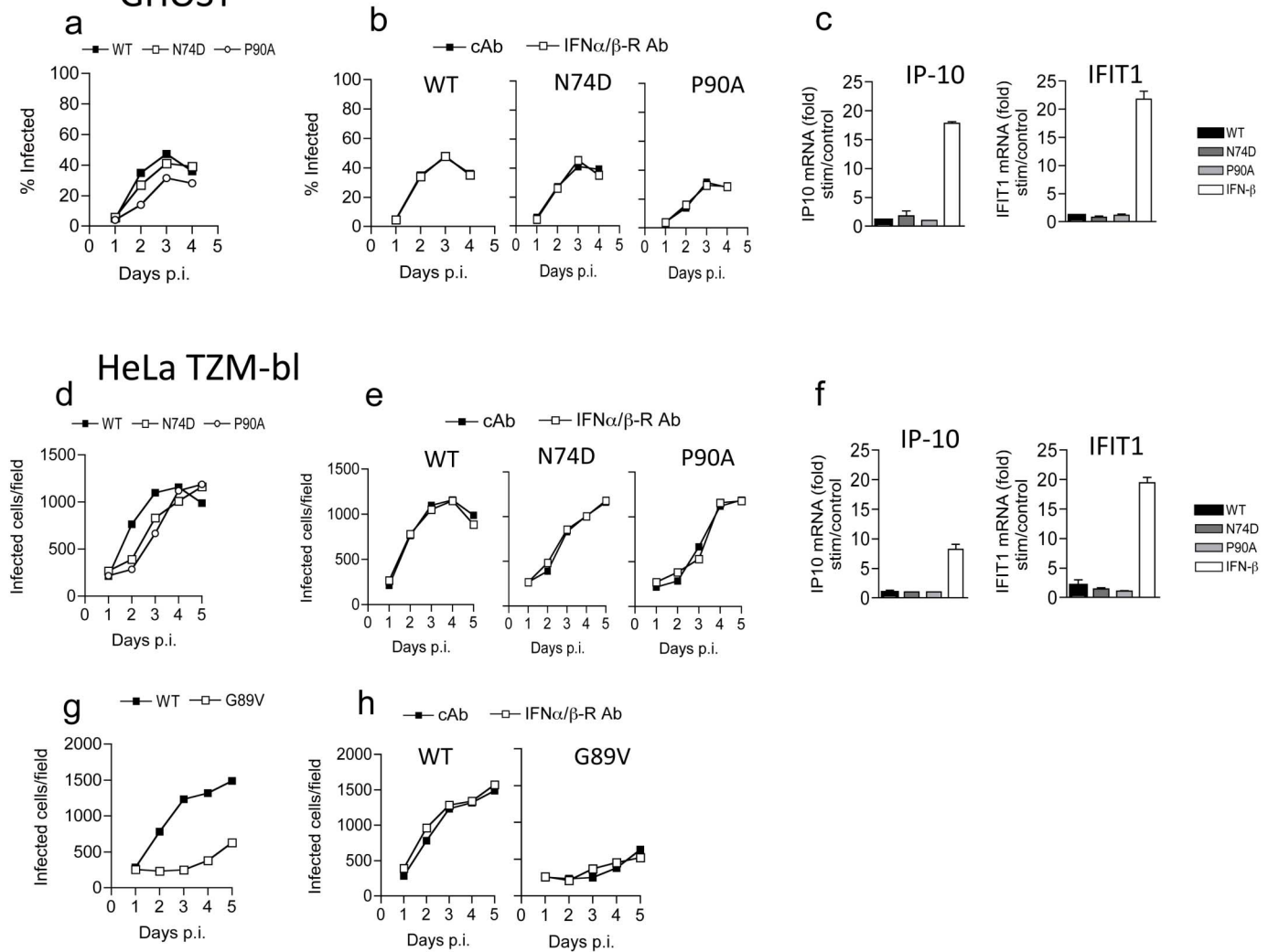
in methods. **b**, Data for 500 single cell measurements are shown for unstimulated MDM and MDM stimulated for 2 h with lipopolysaccharide (LPS) (100 ng ml^{-1}), or infected with wild-type (WT) HIV-1 or N74D or P90A CA mutants. Red lines represent the mean of each data set. * represent statistically significant differences between data sets ($P < 0.01$, t -test), ns represents non-significant differences ($P > 0.05$, t -test).



Extended Data Figure 8 | Inhibition of HIV-1 with cyclosporine or by TREX depletion triggers IFN production. **a**, Infection of human MDM with WT NL4.3 (Ba-L Env) in the presence or absence of 5 μ M cyclosporine. Cells were stained for p24 at specific time points after infection and infected colonies counted. **b**, Supernatants isolated from MDMs in **a** were assayed for soluble IFN- β levels by ELISA. **c**, MDM were infected at low multiplicity with WT HIV in the presence of cyclosporine and IFNAR2 antibody or isotype cAb. **d**, Infectious titre of WT HIV was determined on MDM in the presence of DMSO or cyclosporine at 48h post infection. The data are presented as infectious units (iu) per nanogram (ng) of reverse transcriptase (RT) measured by ELISA. **e**, To confirm that SmBz-CsA inhibits recruitment of cyclophilin A (CypA), but not Nup358 Cyp, to HIV-1 CA, we used the TRIMCyp restriction assay. HIV-1 GFP vector titer on CRFK cells expressing empty vector (EV), HA-tagged owl monkey TRIMCyp RBCC domain fused to human CypA (TRIMCypA) or human Nup358Cyp (TRIMNup358) in either DMSO, or 5 μ M cyclosporine or 10 μ M SmBz-CsA. Protein levels were measured by immunoblot detecting the HA tag with β -Actin as a loading control. In this

assay both cyclosporine and SmBz-CsA inhibited CypA recruitment to CA and rescued VSV-G pseudotyped HIV-1 GFP infectivity from restriction by TRIMCypA. However, neither drug rescued HIV-1 infectivity from TRIMNup358, confirming cyclosporine specificity for CypA and not Nup358. **f**, TREX-1 expression was determined in MDM expressing TREX-1 specific shRNA or control shRNA by qRT-PCR, normalized to GAPDH at the time of HIV-1 infection in **g**. **g**, Cells were stained for p24 at specific time points after WT HIV-1 infection of TREX-1 depleted and control shRNA expressing MDM and infected colonies counted. **h**, Supernatants isolated from MDM in **g** were assayed for soluble IFN- β by ELISA. **i**, Infection of TREX-1 depleted MDM with WT HIV-1 at low multiplicity in the presence of either IFNAR2 antibody or isotype cAb. Cells were stained for p24 at specific time points after infection and infected colonies counted. **j**, Hairpin transduced MDM were assayed for the production of soluble IFN- β levels before HIV-1 infection. Data are representative of 3 independent biological replicates. For HIV-1 replication assays in MDM data points and nonlinear regression lines over time are shown.

GHOST



Extended Data Figure 9 | HIV-1 CA mutants replicate without triggering IFN production in cell lines. GHOST (a–c) or HeLa TzM bl (d–f) indicator cell lines were infected with WT NL4.3 (Ba-L Env) or NL4.3 (Ba-L Env) bearing CA mutations N74D or P90A at low multiplicity (0.04). Replication was monitored by GFP expression (GHOST) or staining LacZ positive cells (HeLa TzM bl). Both mutants replicated well, slightly behind wild-type virus. Replication in GHOST (b) or HeLa TzM bl (e) was performed in the presence of IFNAR2 antibody or isotype cAb. Neither antibody had any effect on WT or mutant HIV-1 replication in GHOST or HeLa TzM bl indicator cell lines. c, f, Induction of ISGs IP10 and IFIT1 expression were measured by

quantitative RT PCR after high multiplicity infection (MOI 2) by WT or CA mutant HIV-1 on GHOST (c) or HeLa TzM bl (f). 10 ng ml^{-1} of IFN- β treatment acted as a positive control. ISG expression levels were normalized to GAPDH and are expressed as fold change in expression over unstimulated cells (Mean of 3 replicates \pm s.e.m.). Neither WT or CA mutant HIV-1 induced ISG expression in either cell line. g, HeLa TzM bl were infected WT or CA G89V at low multiplicity (0.04). Replication was monitored by staining LacZ positive cells. h, HIV-1 CA G89V replication in HeLa TzM bl was defective and not rescued by anti-IFNAR2 or isotype control antibodies. All results are representative of 2 biological replicates.

Extended Data Table 1 | Structure data collection and refinement statistics

Extended Data Table 1 Structure data collection and refinement statistics

CypA:SmBz-CsA	
Data collection	
Space group	I222
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	54.18, 64.60, 80.07
<i>a</i> , <i>b</i> , <i>c</i> (°)	90.00, 90.00, 90.00
Resolution (Å)	32.30-1.67 (1.76-1.67) *
<i>R</i> _{merge}	0.061 (0.562)
<i>I</i> / <i>σ</i>	17.1 (2.9)
Completeness (%)	98.2 (91.1)
Redundancy	6.0 (5.8)
Refinement	
Resolution (Å)	1.67
No. reflections	15648
<i>R</i> _{work} / <i>R</i> _{free}	0.172/0.218
No. atoms	
Protein	1250
Ligand/ion	97
Water	184
B-factors	
Protein	19.485
Ligand/ion	25.523
Water	30.397
R.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	1.082

*Highest resolution shell is shown in parenthesis.

Antigen-specific B-cell receptor sensitizes B cells to infection by influenza virus

Stephanie K. Dougan^{1*}, Joseph Ashour^{1*}, Roos A. Karssemeijer¹, Maximilian W. Popp^{1,2}, Ana M. Avalos¹, Marta Barisa¹, Arwen F. Altenburg¹, Jessica R. Ingram¹, Juan Jose Cragolini¹, Chunguang Guo³, Frederick W. Alt³, Rudolf Jaenisch¹ & Hidde L. Ploegh^{1,2}

Influenza A virus-specific B lymphocytes and the antibodies they produce protect against infection¹. However, the outcome of interactions between an influenza haemagglutinin-specific B cell via its receptor (BCR) and virus is unclear. Through somatic cell nuclear transfer we generated mice that harbour B cells with a BCR specific for the haemagglutinin of influenza A/WSN/33 virus (FluBI mice). Their B cells secrete an immunoglobulin gamma 2b that neutralizes infectious virus. Whereas B cells from FluBI and control mice bind equivalent amounts of virus through interaction of haemagglutinin with surface-disposed sialic acids, the A/WSN/33 virus infects only the haemagglutinin-specific B cells. Mere binding of virus is not sufficient for infection of B cells: this requires interactions of the BCR with haemagglutinin, causing both disruption of antibody secretion and FluBI B-cell death within 18 h. In mice infected with A/WSN/33, lung-resident FluBI B cells are infected by the virus, thus delaying the onset of protective antibody release into the lungs, whereas FluBI cells in the draining lymph node are not infected and proliferate. We propose that influenza targets and kills influenza-specific B cells in the lung, thus allowing the virus to gain purchase before the initiation of an effective adaptive response.

Memory B lymphocytes contribute to the protective immune response to flu infection by producing immunoglobulins that bind and neutralize the virus¹. The lung of an exposed individual contains influenza-specific memory B cells that bind virus, differentiate into plasma cells and secrete either immunoglobulin G (IgG) or IgA locally, reducing the spread of virus^{2,3}. However, the fate of virus-specific B cells that encounter live influenza virus remains unknown.

The low frequency of antigen-specific B cells has hampered analysis of the interactions between live virus, flu antigens and the primary B cells specific for them². To detect influenza-virus-specific B cells, we used sortase-mediated labelling to install Alexa 647 fluorophore onto the haemagglutinin (HA) protein^{4–6}. Virus was disrupted with detergent, HA–Alexa 647 was purified by immunoprecipitation and dialysed to form fluorescent flu micelles (Extended Data Fig. 1a–d). These flu micelles did not stain splenocytes from uninfected mice, but did stain a small number of CD19⁺ cells in spleens of mice infected with influenza and boosted multiple times with A/WSN/33 in incomplete Freund's adjuvant (Extended Data Fig. 1e).

Virus-specific CD19⁺ B cells, isolated by fluorescence-activated cell sorting (Fig. 1a), were used as a source of nuclei for somatic cell nuclear transfer^{7–9}. We transferred the nuclei of these B cells into enucleated oocytes and derived embryonic stem cells^{7–9} that harbour the VDJ/VJ (heavy chain/light chain) rearrangements of the original donor B cell to produce chimaeric mice. We screened offspring of the founder chimaeras by ELISA for the presence of anti-flu antibodies and obtained one animal that showed high titres of IgG2b flu-specific antibodies (Extended Data Fig. 2) and IgG2b⁺IgM[−] B cells in the absence of infection (Fig. 1b). We backcrossed this mouse to C57BL/6 to secure

germline transmission of the VDJ/VJ pair and hereafter refer to the line as FluBI. We know of no other mouse model that harbours B cells of known pathogen specificity or whose primary B cells produce IgG2b.

The sequences of the rearranged heavy and light chain genes (Extended Data Fig. 3) show 7 and 4 somatic mutations in the V_H and V_K segments, respectively. We established the specificity of the FluBI IgG2b antibody

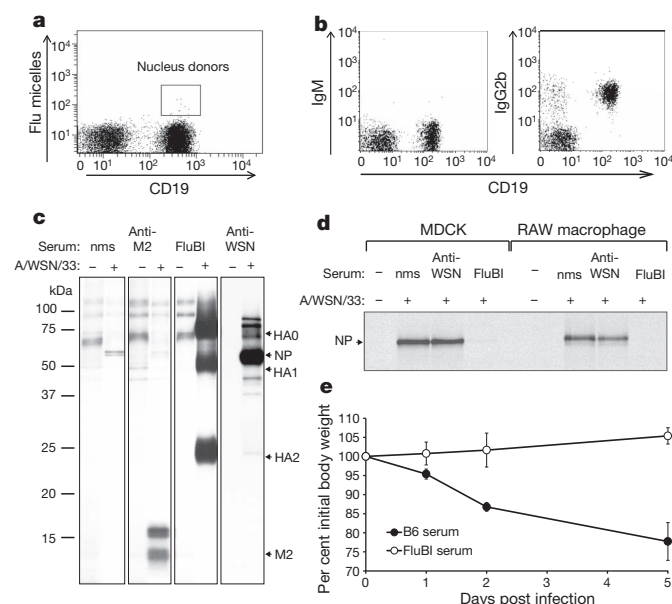


Figure 1 | FluBI mice obtained by somatic cell nuclear transfer from the nucleus of an HA-specific IgG2b⁺ B cell. **a**, A B6x129F1 mouse was infected intranasally with A/WSN/33 and immunized intraperitoneally at days 7, 14 and 21 post-infection with disrupted A/WSN/33 in incomplete Freund's adjuvant. Splenocytes were collected at day 28 post-infection, and stained with anti-CD19 and Alexa 647-labelled HA flu glycoprotein micelles. Cells indicated were sorted and used as donor nuclei for somatic cell nuclear transfer. **b**, Peripheral blood from FluBI mice with germline transmission of the rearranged VDJ and VJ genes was stained with the indicated antibodies and analysed by cytofluorometry. **c**, MDCK cells were infected with A/WSN/33 and labelled with [³⁵S]-cysteine/methionine for 4 h before lysis. Lysates were immunoprecipitated with monoclonal anti-M2, FluBI serum (FluBI), or serum from uninfected (nms) or A/WSN/33-infected mice (Anti-WSN). Immunoprecipitates were analysed by SDS-PAGE and autoradiography. All panels were from the same gel; anti-WSN panels shown are from a shorter exposure time. **d**, A/WSN/33 virus was incubated with the indicated serums before infection of MDCK cells or RAW macrophages. At 2 h.p.i., cells were labelled with [³⁵S]-cysteine/methionine for 2 h, lysed and immunoprecipitated with anti-WSN serum. **e**, BALB/c mice received 100 μ l of serum intravenously from wild-type or FluBI mice, before intranasal challenge with A/WSN/33 (2×10^5 plaque-forming units (p.f.u.) per mouse). $n = 5$; error bars, s.d.

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Boston Children's Hospital, Karp Family Research Building, One Blackfan Circle Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

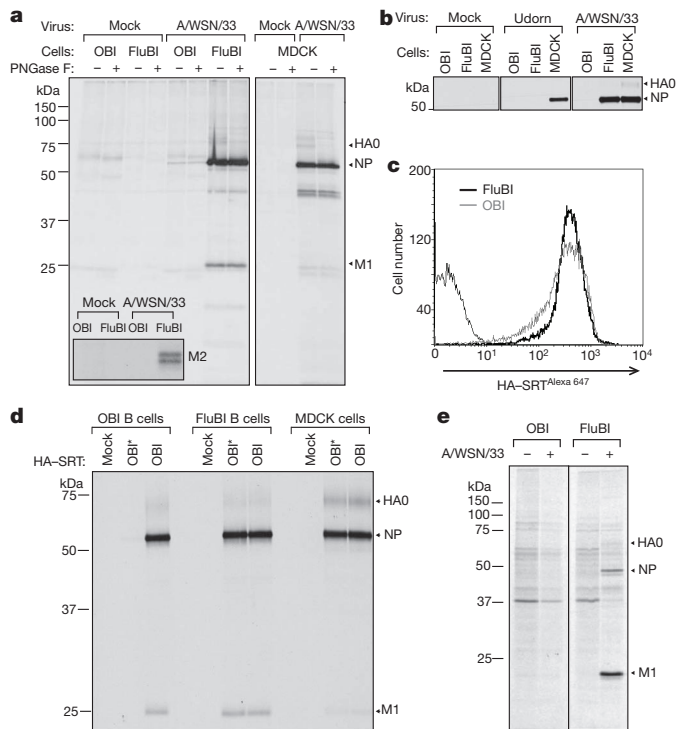


Figure 2 | Influenza virus targets B cells for infection through the BCR.

a, CD40-activated OBI and FluBI B cells and MDCK cells were incubated with A/WSN/33 at a multiplicity of infection (MOI) of 1.0 for 30 min on ice, washed, and transferred to 37 °C in RPMI (0.2% BSA). At 2 h.p.i., cells were labelled with [³⁵S]cysteine/methionine for 2 h, immunoprecipitated with anti-WSN serum or anti-M2 antibody (inset), digested with peptide N-glycosidase F (PNGase F), and analysed by SDS-PAGE and autoradiography. **b**, MDCK cells and CD40-activated OBI or FluBI B cells were incubated with A/WSN/33 or A/Udorn/307/1972 (H3N2) virus and analysed as in **a**. **c**, CD40-activated OBI or FluBI B cells were incubated with HA-SRT^{Alexa 647} virus for 30 min on ice and analysed by cytofluorometry. **d**, CD40-activated OBI or FluBI B cells or MDCK cells were incubated on ice for 30 min with HA-SRT virus modified with a 17-mer peptide containing the OBI epitope (OBI) or a mutant version that no longer binds to OBI (OBI*). Infection was analysed as in **a**. **e**, CD40-activated OBI or FluBI B cells were infected with A/WSN/33 at a MOI of 1.0 and labelled with [³⁵S]cysteine/methionine at 4 h.p.i. for 4 h. Released virus particles were recovered by adsorption to chicken erythrocytes and analysed by SDS-PAGE and autoradiography.

by immunoprecipitation from lysates of [³⁵S]cysteine/methionine-labelled, A/WSN/33-infected MDCK cells (Fig. 1c). The antibody retrieves HA0 and its cleavage products¹⁰ HA1 and HA2. FluBI IgG2b antibody purified from hybridomas generated from FluBI;Rag2^{-/-} splenocytes gave similar results (Methods). The serum from FluBI mice neutralizes A/WSN/33 *in vitro* (Fig. 1d) and *in vivo* (Fig. 1e). Cytofluorometry of B-cell populations in lymph node, spleen and bone marrow from FluBI mice showed a complete absence of B-1a B cells, and other B-cell subsets were near-normal in distribution and number (Extended Data Fig. 4). As shown for OBI mice⁷, the presence of a functionally rearranged $\gamma 2b$ heavy chain locus does not compromise B-cell development, despite the deletion of the μ , δ , $\gamma 3$ and $\gamma 1$ constant regions in FluBI mice.

To determine the fate of HA-specific B cells upon encounter with virus, we obtained B cells from the FluBI mouse and from OBI mice, whose B cells produce an IgG1 specific for ovalbumin⁷. Before infection, we activated cells overnight with anti-CD40 to improve biosynthetic labelling, used to assess viral antigen synthesis. At 2 h post infection (h.p.i.) FluBI B cells infected with A/WSN/33 synthesize vastly more HA, nucleoprotein (NP) and M2 protein (inset) than OBI B cells (Fig. 2a). In FluBI B cells, the levels of NP were comparable to those obtained from A/WSN/33-infected MDCK cells, an indication that replication

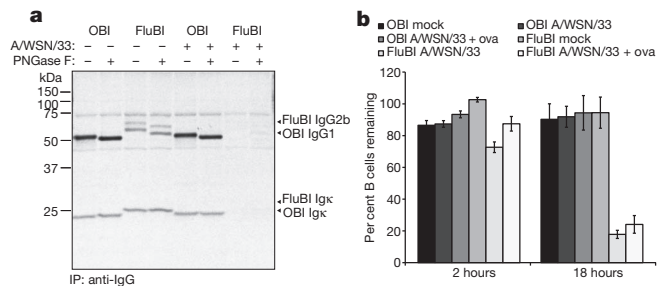


Figure 3 | Influenza-infected B cells fail to secrete antibodies and ultimately die. **a**, CD40-activated OBI or FluBI B cells were incubated with A/WSN/33 for 30 min on ice, washed and cultured in RPMI (0.2% BSA) for 2 h. Cells were labelled with [³⁵S]cysteine/methionine for 2 h, and immunoglobulins were recovered from supernatants using protein-G agarose. IP, immunoprecipitation. **b**, CD40-activated OBI or FluBI B cells were incubated with A/WSN/33 for 30 min on ice, washed and mixed at a 1:1 ratio with mock-infected MHCII-GFP CD40-activated B cells. Ovalbumin (100 μ g ml⁻¹) was included where indicated. Cells were cultured in complete RPMI for 2 or 18 h, stained with anti-CD19 and 7-aminoactinomycin D, and analysed by cytofluorometry. Live cells were calculated as (no. GFP⁺ CD19⁺ cells/no. GFP⁺ CD19⁺ cells)/(no. GFP⁺ CD19⁺ cells/no. GFP⁺ CD19⁺ cells at time 0) \times 100. Error bars are s.d. of triplicate cultures.

of A/WSN/33 in FluBI B cells is robust (Fig. 2a and Extended Data Fig. 5). Neither FluBI nor OBI B cells were infected by the closely related strain A/Puerto Rico/8/1934 (H1N1) or with A/Udorn/307/1972 (H3N2) (Fig. 2b and Extended Data Figs 6, 7).

The increased levels of antigen detected in FluBI B cells might result from improved binding of virus via BCR-HA interactions. To measure virus binding, we incubated anti-CD40-activated B cells with sortase-modifiable (HA-SRT) Alexa 647-labelled virus (HA-SRT^{Alexa 647})⁴ and measured bound virus by cytofluorometry. Virus bound equally well to FluBI and OBI B cells (Fig. 2c). To determine whether increased susceptibility of FluBI B cells to infection is indeed BCR-dependent, we generated HA-SRT virus, transacylated⁴ at the carboxy terminus of HA1 with a synthetic 17-residue ovalbumin peptide that comprises the epitope recognized by the ovalbumin-specific OBI B cells (HA-SRT^{wtOBI}). This virus should now also bind to the BCR expressed on the surface of OBI B cells. For comparison we labelled HA-SRT virus with a mutant version of the OBI 17-mer peptide (HA-SRT^{mtOBI}) no longer recognized by the OBI IgG1. We then exposed FluBI and OBI B cells to either HA-SRT^{wtOBI} or HA-SRT^{mtOBI} virus (Fig. 2d). As expected, the two HA-SRT viruses infected FluBI B cells equally, regardless of the identity of the peptide epitope installed. In contrast, only HA-SRT^{wtOBI} infected OBI B cells. The level of infection in OBI B cells exposed to HA-SRT^{wtOBI} was similar to that seen in FluBI B cells. The presence of a BCR that recognizes HA, native or modified to impart BCR reactivity, thus causes susceptibility to influenza infection. Mere adsorption of virus to the cell surface through interactions with sialic acids may therefore not suffice to gain entry¹¹⁻¹³, and interaction with an internalizing receptor is important¹⁴⁻¹⁷. Antigen-occupied BCRs are indeed efficiently internalized¹⁸, especially when engaged by a multivalent ligand, and could thus improve virus entry and infection.

Do infected B cells produce virus or virus-like particles (VLPs)? We infected OBI and FluBI B cells with A/WSN/33 and biosynthetically labelled them 4–6 h.p.i. We incubated the culture supernatants with chicken erythrocytes to recover released virus and VLPs via interactions of HA with erythrocyte-borne sialic acids, and analysed adsorbed materials by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and autoradiography to detect the presence of viral structural proteins (Fig. 2e). NP and M1 proteins were present in material released by FluBI but not OBI B cells, indicating that infection of FluBI B cells was productive.

We also looked for secreted antibody from A/WSN/33-exposed B cells (Fig. 3a). OBI cells showed no change in their ability to secrete

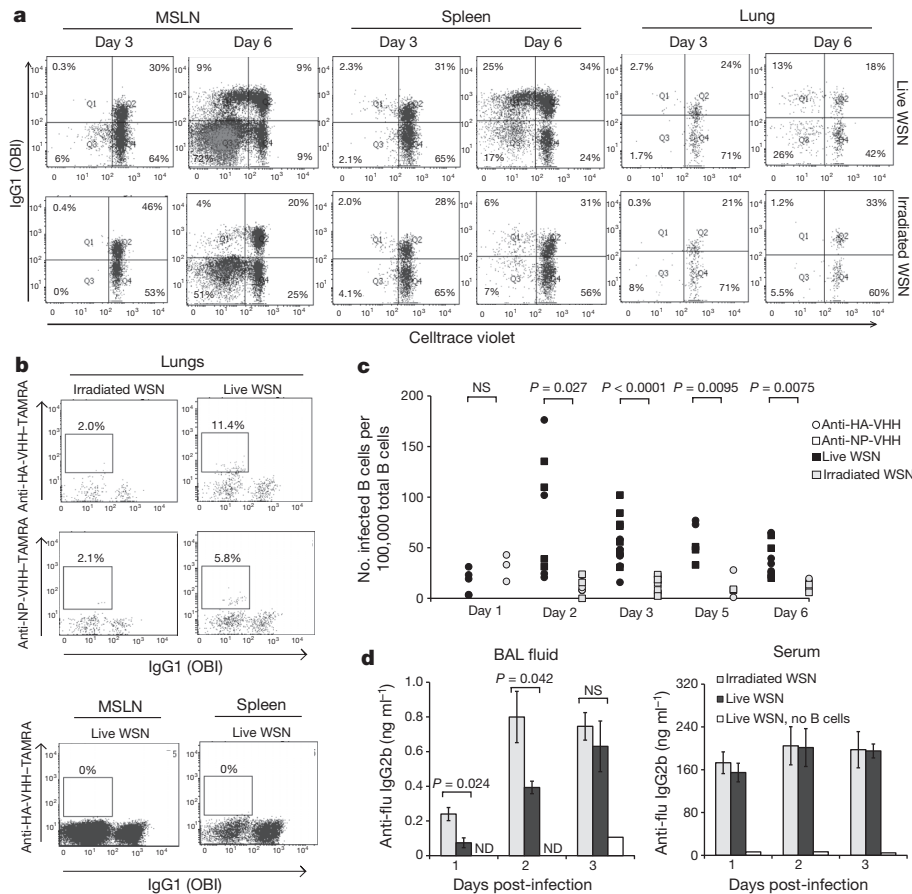


Figure 4 | B cells can be infected with A/WSN/33 in the lungs, but not in the draining lymph node. Naive B cells from OBI;MHCII-GFP and FluBI;MHCII-GFP mice were stained with Celltrace violet, mixed in a 1:1 ratio, and transferred intravenously into C57BL/6 mice (10^7 total cells per recipient). Recipient mice were infected intranasally with live or irradiated A/WSN/33 (2×10^5 p.f.u. per mouse) mixed with ovalbumin (100 μ g per mouse) and euthanized at 3 or 6 days post-infection. Cells were collected from spleen, mediastinal lymph nodes (MSLN) and lungs. **a**, Plots are gated on GFP⁺-transferred cells. Dilution of violet dye indicates proliferation. Anti-IgG1 distinguishes OBI from FluBI B cells. Plots are representative of 4 mice per group. **b**, Cells from **a** were permeabilized, fixed and stained with

antibody after exposure to A/WSN/33, presumably because they are poorly infected, if at all. Not only did FluBI B cells exposed to A/WSN/33 show decreased levels of secreted IgG2b, they also died within 18 h.p.i. (Fig. 3b) under conditions where OBI B cells survived.

To explore the consequences of *in vivo* exposure of FluBI cells to virus and track FluBI B cells, we transferred Celltrace violet-marked OBI (IgG1⁺) and FluBI (IgG1⁻) B cells, obtained from transnuclear animals crossed onto the GFP-tagged major histocompatibility complex class II (MHCII-GFP) background, into naive recipients. The number of B cells transferred was chosen such that it still allowed detection of transferred cells in lungs and mediastinal lymph nodes from infected mice. Mice were inoculated intranasally with live or irradiated A/WSN/33 and mixed with ovalbumin to ensure engagement of the respective donor BCRs. We then measured cell number, proliferation and expression of viral antigens on donor MHCII-GFP⁺ B cells (Fig. 4). The ovalbumin-specific OBI B cells showed robust proliferation in both spleen and mediastinal lymph nodes, whereas FluBI B cells proliferated to a greater extent in the mediastinal lymph nodes in response to live virus, probably owing to the increased amount of antigen present locally following viral replication in the lungs. By day 6 post-infection, proliferating FluBI B cells in the lymph nodes had differentiated into CD138⁺ plasmablasts (Extended Data Fig. 8). Mice that received irradiated A/WSN/33 showed less proliferation of FluBI

single-domain antibodies recognizing HA (anti-HA-VHH) or NP (anti-NP-VHH). Plots are representative of 5 mice per group. **c**, Quantification of viral antigen-positive B cells as shown in **b**. *P* values were determined using a one-sided *t*-test with Bonferroni correction. NS, not significant. **d**, C57BL/6 mice were administered 10^7 FluBI B cells intravenously and infected intranasally with live or ultraviolet-irradiated A/WSN/33. Mice were euthanized at 1, 2 or 3 days post-infection. Bronchoalveolar lavage (BAL) fluid and serum were analysed by ELISA for A/WSN/33 reactivity using horseradish peroxidase-coupled anti-IgG2b. ND, not detected. *P* values were determined using a two-sided *t*-test with Bonferroni correction. Day 1, *n* = 5; day 2, *n* = 7 irradiated, *n* = 8 live; day 3, *n* = 4. Error bars are s.e.m.

cells in mediastinal lymph nodes owing to the absence of antigen synthesis, but more proliferation of FluBI cells in the lungs at 3 days post-infection than animals that received live virus (Fig. 4a; day 3 lungs live $2.7 \pm 0.43\%$ *n* = 6 versus day 3 lungs irradiated $6.5 \pm 1.3\%$ *n* = 5; *P* = 0.015).

To detect intracellular influenza viral proteins as an additional measure of infection, we generated HA- and NP-specific heavy-chain-only antibody fragments (VHHs) from an influenza-immunized alpaca (Extended Data Fig. 9). These small, single-domain VHHs are C-terminally modified with an LPETG motif for direct coupling with TAMRA using sortase¹⁹ and detect flu antigens in FluBI B cells infected *in vitro* (Extended Data Fig. 10). In mice infected with live A/WSN/33 we observed HA- and NP-positive, infected FluBI cells in the lung (Fig. 4b, c). Although virus is reportedly delivered to the mediastinal lymph nodes by dendritic cells^{20,21}, no flu-antigen-positive B cells were found in the mediastinal lymph nodes (Fig. 4b). Not unexpectedly, mice exposed to irradiated A/WSN/33 lacked flu-antigen-positive B cells in either location. Co-transferred OBI cells were present in the lungs of infected mice, but were HA- and NP-negative.

To determine whether infection of lung-resident flu-specific B cells affects antibody production *in vivo*, we measured in bronchoalveolar lavage (BAL) fluid the presence of flu-specific IgG2b, secreted by the transferred FluBI B cells, and found that irradiated virus elicited a

stronger and more rapid initial response than live virus, consistent with the ability of live, but not irradiated virus to kill FluBI B cells. Serum levels of flu-specific IgG2b were equivalent in mice receiving live versus irradiated virus, indicating that loss of flu-specific IgG2b is restricted to the lung bronchoalveolar space.

FluBI cells, specific for haemagglutinin and secreting a neutralizing antibody, themselves succumb to infection mediated by the surface-disposed BCR. The rapid death of A/WSN/33-specific FluBI cells provides respite for the virus at the lung epithelium, a site to which antigen-specific B cells are recruited in the course of infection and where they remain as sentinels thereafter^{2,22,23}. Infection and killing of a fraction of the rare antigen-specific B cells impairs the kinetics of the memory response, and confers an advantage to the virus with its replication cycle measured in hours. The ability of a pathogen to exploit this mode of entry and eliminate the initial wave of the very B cells capable of counteracting the infection is an efficient means of ensuring a window for replication and horizontal transmission. It is unlikely to be limited to influenza virus.

METHODS SUMMARY

Sortase labelling. HA-SRT virus (a derivative of A/WSN/33, ref. 6) was incubated with Sortase A (150 μ M) and the indicated nucleophile (500 μ M) in sortase labelling buffer with 0.2% BSA at 37 °C for 1 h, resulting in site-specific labelling of HA on the intact virus particle. Labelled virus was concentrated over a 20% sucrose cushion. For flu micelles, HA-SRT^{Alexa 647} was disrupted with detergent, immunoprecipitated with anti-Alexa 647 and dialysed to form HA-enriched micelles. For OBI epitope labelling, HA-SRT was labelled with OBI (GGGFDKLPFGDSI EAQGGK) or OBI* (GGGFDKLPAGASIEAQGGK).

Hybridoma production. FluBI;Rag2^{-/-} spleen cells were fused with NS-1 cells. The resulting hybridomas were screened for A/WSN/33 reactivity by ELISA using horseradish peroxidase-coupled anti-IgG2b.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 May; accepted 4 September 2013.

Published online 20 October 2013.

- Valkenburg, S. A. *et al.* Immunity to seasonal and pandemic influenza A viruses. *Microbes Infect.* **13**, 489–501 (2011).
- Onodera, T. *et al.* Memory B cells in the lung participate in protective humoral immune responses to pulmonary influenza virus reinfection. *Proc. Natl Acad. Sci. USA* **109**, 2485–2490 (2012).
- Manicassamy, B. *et al.* Analysis of *in vivo* dynamics of influenza virus infection in mice using a GFP reporter virus. *Proc. Natl Acad. Sci. USA* **107**, 11531–11536 (2010).
- Popp, M. W., Karssemeijer, R. A. & Ploegh, H. L. Chemoenzymatic site-specific labeling of influenza glycoproteins as a tool to observe virus budding in real time. *PLoS Pathog.* **8**, e1002604 (2012).
- Popp, M. W. & Ploegh, H. L. Making and breaking peptide bonds: protein engineering using sortase. *Angew. Chem. Int. Edn Engl.* **50**, 5024–5032 (2011).
- Antos, J. M., Miller, G. M., Grotenbreg, G. M. & Ploegh, H. L. Lipid modification of proteins through sortase-catalyzed transpeptidation. *J. Am. Chem. Soc.* **130**, 16338–16343 (2008).
- Dougan, S. K. *et al.* IgG1⁺ ovalbumin-specific B-cell transnuclear mice show class switch recombination in rare allelically included B cells. *Proc. Natl Acad. Sci. USA* **109**, 13739–13744 (2012).
- Kirak, O. *et al.* Transnuclear mice with predefined T cell receptor specificities against *Toxoplasma gondii* obtained via SCNT. *Science* **328**, 243–248 (2010).
- Dougan, S. K. *et al.* Transnuclear TRP1-specific CD8 T cells with high or low affinity TCRs show equivalent anti-tumor activity. *Cancer Immunol. Res.* **1**, 99–111 (2013).
- Wiley, D. C. & Skehel, J. J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* **56**, 365–394 (1987).
- Stray, S. J., Cummings, R. D. & Air, G. M. Influenza virus infection of desialylated cells. *Glycobiology* **10**, 649–658 (2000).
- Thompson, C. I., Barclay, W. S., Zambon, M. C. & Pickles, R. J. Infection of human airway epithelium by human and avian strains of influenza A virus. *J. Virol.* **80**, 8060–8068 (2006).
- Pedroso de Lima, M. C. *et al.* Target cell membrane sialic acid modulates both binding and fusion activity of influenza virus. *Biochim. Biophys. Acta* **1236**, 323–330 (1995).
- Huang, R. T., Lichtenberg, B. & Rick, O. Involvement of annexin V in the entry of influenza viruses and role of phospholipids in infection. *FEBS Lett.* **392**, 59–62 (1996).
- Chu, V. C. & Whittaker, G. R. Influenza virus entry and infection require host cell N-linked glycoprotein. *Proc. Natl Acad. Sci. USA* **101**, 18153–18158 (2004).
- Londrigan, S. L. *et al.* N-linked glycosylation facilitates sialic acid-independent attachment and entry of influenza A viruses into cells expressing DC-SIGN or L-SIGN. *J. Virol.* **85**, 2990–3000 (2011).
- Eierhoff, T., Hrinčius, E. R., Rescher, U., Ludwig, S. & Ehrhardt, C. The epidermal growth factor receptor (EGFR) promotes uptake of influenza A viruses (IAV) into host cells. *PLoS Pathog.* **6**, e1001099 (2010).
- Harwood, N. E. & Batista, F. D. Early events in B cell activation. *Annu. Rev. Immunol.* **28**, 185–210 (2010).
- Witte, M. D. *et al.* Preparation of unnatural N-to-N and C-to-C protein fusions. *Proc. Natl Acad. Sci. USA* **109**, 11993–11998 (2012).
- Moldedo, B. *et al.* Cutting edge: stealth influenza virus replication precedes the initiation of adaptive immunity. *J. Immunol.* **183**, 3569–3573 (2009).
- Moldedo, B., Li, W., Yount, J. S. & Moran, T. M. Unique type I interferon responses determine the functional fate of migratory lung dendritic cells during influenza virus infection. *PLoS Pathog.* **7**, e1002345 (2011).
- Joo, H. M., He, Y. & Sangster, M. Y. Broad dispersion and lung localization of virus-specific memory B cells induced by influenza pneumonia. *Proc. Natl Acad. Sci. USA* **105**, 3485–3490 (2008).
- Jones, P. D. & Ada, G. L. Influenza-specific antibody-secreting cells and B cell memory in the murine lung after immunization with wild-type, cold-adapted variant and inactivated influenza viruses. *Vaccine* **5**, 244–248 (1987).

Acknowledgements S.K.D. and C.G. were funded by the Cancer Research Institute. J.J.C. was funded by the Human Frontiers Science Program. F.W.A., R.J. and H.L.P. are funded by grants from the National Institutes of Health. F.W.A. is a Howard Hughes Medical Institute investigator. S.K.D. and H.L.P. are funded by the American Association for Cancer Research-Pancreatic Cancer Action Network. We are grateful to P. Wisniewski for cell sorting, to J. Jackson for mouse husbandry, to G. Bell for statistical analysis, to N. Watson for electron microscopy and to M. Witte for sortase nucleophiles.

Author Contributions S.K.D. and J.A. contributed equally. Somatic cell nuclear transfer was performed by S.K.D.; S.K.D., J.A., R.A.K., M.W.P., M.B. and A.F.A. performed experiments. A.M.A. generated hybridomas. J.R.I. and J.J.C. generated flu-specific VHs. C.G. sequenced the BCR loci. F.W.A. and R.J. provided advice and reagents. S.K.D., J.A. and H.L.P. designed experiments, analysed data and wrote the paper.

Author Information Sequences of the mouse *V_D* and *J* genes were deposited in GenBank with accession numbers KF419287 and KF419288. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.L.P. (ploegh@wi.mit.edu).

METHODS

Reagents. Anti A/WSN/33 serum was generated by infecting BALB/C mice with A/WSN/33 (2×10^5 p.f.u. per mouse). Anti-M2 antibody (14C2) was purchased from Santa Cruz Biotechnology. OBI peptides were provided by the MIT biopolymers facility. The amino acid sequence that comprises the OBI epitope is as follows: GGGFDKLPFGDSEIAQGQK. The mutant sequence that fails to bind to the OBI antibody is as follows (substitutions denoted in bold): GGGFDKLPAGASIEAQGGK. Sortase was expressed in *Escherichia coli* and purified as described²⁴. The GGGK-Alexa 647 peptide used to label HA-SRT was provided by M. Witte. Anti IgG-FITC antibody (A1101) was purchased from Molecular Probes. Chicken erythrocytes (cRBCs) were purchased from Lampire Biological Laboratories. Express³⁵S protein labelling mix was purchased from Perkin Elmer. Methionine and cysteine-free RPMI, Optimum, HEPES buffer and non-essential amino acids (NEAA) were purchased from Life Technologies. Endoglycosidase H (EndoH) and PNGase F were purchased from New England Biolabs. Protein-G agarose and EDTA-free complete protease inhibitor cocktail was purchased from Roche Diagnostics.

Virus propagation and infection. A/Puerto Rico/8/1934 (H1N1) and A/Udorn/307/1972 (H3N2) viral stocks were a gift from X. Zhuang. MDCK cells and RAW cells were originally obtained from the ATCC, and are tested for mycoplasma every 3–6 months. A/WSN/33 and HA-SRT virus⁴ were propagated in MDCK cells grown in Optimum and supplemented with $1 \mu\text{g ml}^{-1}$ L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK)-treated trypsin. Titres of A/WSN/33 and HA-SRT virus stocks were determined by plaque assay on MDCK cells. For the plaque assay, MDCK cells were cultured in 24-well dishes until sub-confluent. Cells were washed twice in PBS supplemented with Ca^{2+} and Mg^{2+} (PBS+) and infected with tenfold serial dilutions of virus in PBS+ supplemented with 0.25% BSA for 1 h at room temperature. Cells were washed once in PBS+ then overlaid with plaque media ($1 \times \text{MEM}$, 0.25% BSA, 0.8% agar, $0.5 \mu\text{g ml}^{-1}$ trypsin-TPCK) and placed at 37°C . After 24 to 48 h, the agar overlay was removed and the cells were fixed with 3% paraformaldehyde and permeabilized using PBS 0.5% NP-40. Influenza plaques were stained using monoclonal antibody against NP conjugated to FITC then visualized and quantified by fluorescent microscopy. Labelled HA-SRT virus was quantified using haemagglutination assay against a standard containing a known quantity of A/WSN/33. For all *in vitro* experimental infections virus was diluted in PBS+/BSA(0.25%) and supplemented with $1 \mu\text{g ml}^{-1}$ TPCK-treated trypsin. In the case of MDCK cells, cells were trypsinized and infections carried out in suspension. Virus and cells were incubated on ice for 30 min. Cells were washed with PBS+, and resuspended in DMEM with 0.2% BSA, 100 mM HEPES and NEAA (or RPMI with 0.2% BSA, 100 mM HEPES and NEAA in experiments where B cells were used).

Sortase labelling of HA-SRT virus. HA-SRT virus⁴ was incubated with sortase A ($150 \mu\text{M}$) and the indicated nucleophile ($500 \mu\text{M}$) in sortase labelling buffer (100 mM Tris pH 7.4, 150 mM NaCl, 10 mM Ca^{2+}) supplemented with 0.2% BSA at 37°C for 1 h. The labelled virus was then concentrated over a 20% sucrose cushion and resuspended in sortase labelling buffer supplemented with 0.2% BSA.

Neutralization assay. 10^5 p.f.u. A/WSN/33 was incubated with 20 μl of the indicated test serum for 30 min on ice. The virus/serum mixture was then overlaid onto 10^6 MDCK or RAW macrophages and incubated for 30 min at room temperature. Cells were washed and incubated at 37°C in DMEM (0.2% BSA). At 2 h.p.i., both cell types were biosynthetically labelled for two hours and subsequently lysed. Immunoprecipitates were prepared with anti-WSN serum and analysed by SDS-PAGE and autoradiography.

Animal care. All mice were housed at the Whitehead Institute for Biomedical Research and were maintained according to protocols approved by the MIT Committee on Animal Care. A/WSN/33-infected animals were housed in an approved quarantine room at Whitehead Institute. C57BL/6 and BALB/c mice were purchased from Jackson Labs. *Rag2*^{-/-} mice (RAGTN12) were purchased from Taconic. MHCII-GFP mice²⁵ and OBI mice⁷ have been described previously.

In vivo influenza infections. 8–10 week old male mice were anesthetized by a single dose of avertin (1.25% tribromoethanol, GIBCO), 250 mg kg⁻¹ body weight, delivered by intraperitoneal injection. A/WSN/33 (2×10^5 p.f.u. per mouse unless otherwise stated) or A/Puerto Rico/8/1934 (2×10^5 p.f.u.) was administered intranasally. Infected BALB/c mice were weighed daily, and mice losing more than 20% of starting weight were euthanized. Infected C57BL/6 mice are intrinsically more resistant to influenza infection²⁶, and did not show more than 5% weight loss; C57BL/6 mice used as hosts for adoptive transfer (intravenously) of FluBI or OBI B cells were infected with A/WSN/33 within 24 h of donor cell transfer. Mice that received adoptive transfer of FluBI and/or OBI B cells were allowed to intermix freely in a large cage before separating mice into groups to receive irradiated or live flu virus. Infections were performed by a non-blinded investigator. Successfully administration of intranasal virus was verified by

recruitment of FluBI cells to the MSLN. <5% of mice showed failure of intranasal delivery and were excluded from analysis, as per pre-established criteria.

Transnuclear mouse generation. Transnuclear mice were generated as previously described^{27–29,28}. Briefly, flu-specific B cells were sorted by FACS and used as a source of donor nuclei for somatic cell nuclear transfer according to the protocol established by the Wakayama group²⁹. The mitotic spindle was removed from mouse oocytes and replaced with donor nuclei. The nucleus-transplanted oocytes were then activated in medium containing strontium and trichostatin A, and allowed to develop in culture to the blastocyst stage. Because the live birth rate of somatic cell nuclear transfer blastocysts transferred into pseudopregnant females is quite low, somatic cell nuclear transfer blastocysts were used instead to derive embryonic stem (ES) cell lines. These ES cell lines were then injected into wild type B6xDBA F1 blastocysts and implanted into pseudopregnant females. The resulting chimaeric pups were mated to C57BL/6 females to establish the FluBI line. All animals used were backcrossed 4–5 generations onto C57BL/6 or C57BL/6;*Rag2*^{-/-} backgrounds.

Flu micelles. HA-SRT virus was incubated with Sortase A ($150 \mu\text{M}$) and Alexa 647 nucleophile ($500 \mu\text{M}$) in sortase labelling buffer (100 mM Tris pH 7.4, 150 mM NaCl, 10 mM Ca^{2+}) supplemented with 0.2% BSA at 37°C for 1 h. HA-SRT^{Alexa 647} virus was disrupted with Triton X100 and incubated with 400 μg anti-Alexa 647 overnight. HA-Alexa 647 was then recovered using protein G-Sepharose. Bound proteins were eluted with 0.1 M glycine pH 2.8. Detergent was dialysed to form protein micelles³⁰.

Flow cytometry. Cells were collected from spleen, pooled mesenteric and cervical lymph nodes, mediastinal lymph nodes, lung, peritoneal cavity and bone marrow. Lung tissue was digested in RPMI with 1% (w/v) collagenase D (Sigma) for 30–60 min at 37°C before mechanical dissociation with a 40- μm cell strainer. Cell preparations were subjected to hypotonic lysis to remove erythrocytes, stained and analysed using a FACS Fortessa (BD). Celltrace violet was purchased from Invitrogen. All antibodies were from BD Pharmingen. For intracellular staining, cells were fixed and permeabilized using Cytofix/Cytoperm (BD) according to the manufacturer's instructions.

Flu-specific ELISAs. High-binding 96-well microtitre plates (Costar) were coated overnight at 4°C with A/WSN/33 (2×10^4 p.f.u. ml⁻¹) in PBS. Plates were washed 3 times with wash buffer (PBS, 0.05% Tween-20), blocked with 10% fetal bovine serum for 1 h at room temperature, washed 3 times, and incubated with samples. Bronchoalveolar lavage (BAL) fluid samples were collected by inserting a 24 gauge catheter into an incision in the trachea, filling the lungs with 1 ml PBS and recovering 0.7–0.8 ml of lavage fluid. BAL fluid samples were used neat. Serum samples were used at 1:10 dilution. FluBI antibody purified from hybridoma supernatants was used as standard. Plates were incubated with samples at room temperature for 2 h, washed 5 times, and incubated with HRP-coupled anti-IgG2b secondary reagent for 1 h. Plates were washed 7 times, and detected using 3,3',5,5'-tetramethylbenzidine (TMB) liquid substrate (Sigma).

Cell culture. B cells were purified from pooled spleen and lymph nodes by negative selection using anti-CD43 magnetic beads (Miltenyi Biotec). B cells were cultured in RPMI 1640 medium supplemented with 10% heat-inactivated FBS, 2 mM L-glutamine, 100 U ml⁻¹ penicillin G sodium, 100 $\mu\text{g ml}^{-1}$ streptomycin sulphate, 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, and 0.1 mM 2-mercaptoethanol. For differentiation of B cells into plasmablasts, anti-CD40 (BD clone HM40-3, $1 \mu\text{g ml}^{-1}$) was added to the culture medium. MDCK cells and RAW macrophages were cultured in DMEM supplemented with 10%FCS, 100 mM HEPES and NEAA. For experiments where B cells and MDCK cells were compared side by side, MDCK cells were cultured during the experiment using RPMI media.

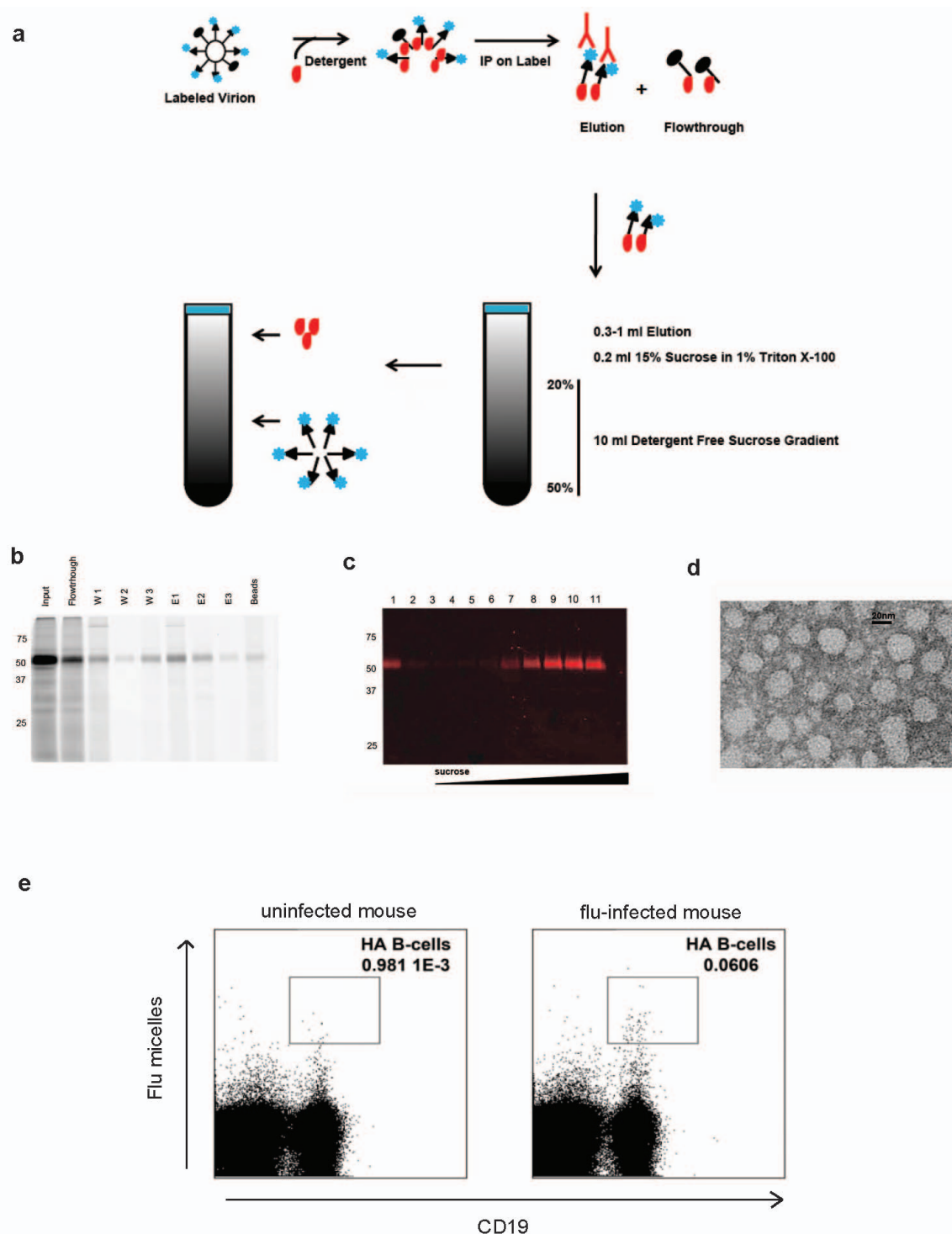
Biosynthetic labelling and immunoprecipitation. Cells were starved for 15 min in DMEM (Cys-, Met-) or RPMI (Cys-, Met-), supplemented with 10% FCS and then labelled with [³⁵S]methionine/cysteine (276 μCi of Express³⁵S protein labelling mix per 10^6 cells. Cells were washed and lysed in NP40 lysis buffer (25 mM Tris pH 7.4, 150 mM NaCl, 5 mM MgCl_2 , 0.5% NP-40). For immunoprecipitation, lysates were incubated with 20 μl of protein-G agarose and the antibody stated or immune serum. Monoclonal anti M2 antibody was used at 2.5 $\mu\text{g ml}^{-1}$. For all other immunoprecipitations, we used 1 μl of serum per ml of lysate. Incubations with antisera were performed for 3 h at 4°C . The protein G Sepharose beads were washed three times in lysis buffer resuspended in SDS sample buffer. Where glycosidase treatment was required, digestions were performed following manufacturer's instructions (New England Biolabs), before addition of SDS sample buffer. To recover released virions form culture supernatants, the amount of supernatant used was normalized to the amount of radioactivity incorporated into cells. The adjusted volumes of supernatant were incubated with chicken red blood cells at a 1:50 dilution for three hours at 4°C . The erythrocytes were washed by centrifugation ($3 \times$) with PBS and lysed in NP40-containing lysis

buffer. Immunoprecipitates and materials adsorbed onto chicken red blood cells were analysed by SDS-PAGE and autoradiography.

Production of FluBI hybridoma. Spleen cells from FluBI;*Rag2*^{-/-} mice were stimulated with 40 µg ml⁻¹ LPS and 20 ng ml⁻¹ IL4 for 5 days, and were fused with NSObcl2 cells (a gift from B. Diamond) and selected in medium supplemented with 20% heat-inactivated FCS and hypoxanthine aminopterin thymidine (HAT) and grown in 10% CO₂ for 3 weeks, before transfer of positive clones to hypoxanthine thymidine (HT) supplemented medium containing 10% heat-inactivated FCS. Resulting hybridomas were screened for A/WSN/33 reactivity by ELISA using HRP-coupled anti-IgG2b or anti-Igκ (Southern Biotech) secondary antibodies for detection.

Statistics. Centre values are mean. *P* < 0.05 defined as significant. Standard two-sided *t*-test was used throughout unless otherwise noted. Sample size was based on variability from pilot studies.

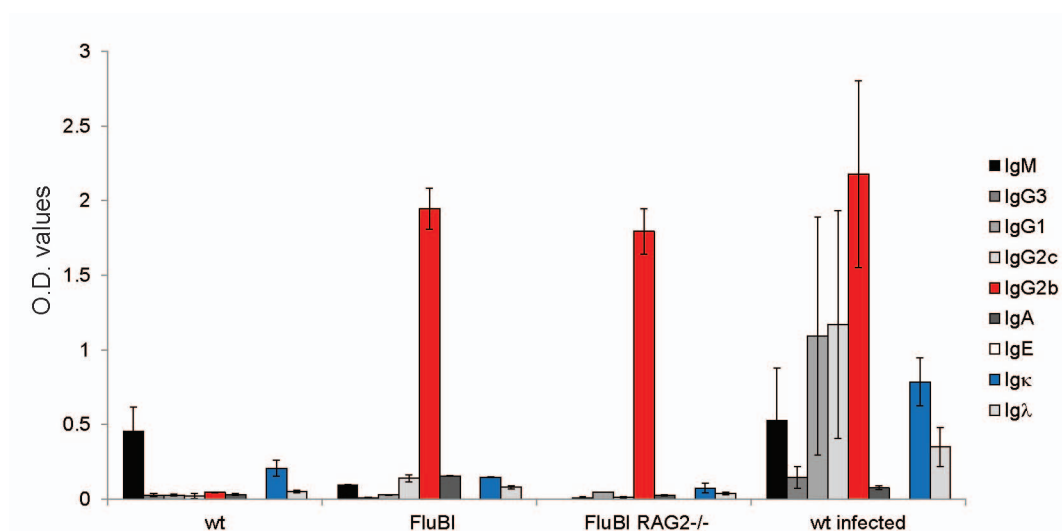
24. Popp, M. W., Antos, J. M. & Ploegh, H. L. Site-specific protein labeling via sortase-mediated transpeptidation. *Curr. Protoc. Protein Sci.* **15**, Unit 15.13 (2009).
25. Boes, M. *et al.* T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. *Nature* **418**, 983–988 (2002).
26. Morgan, D. J., McLain, L. & Dimmock, N. J. Protection of three strains of mice against lethal influenza in vivo by defective interfering virus. *Virus Res.* **29**, 179–193 (1993).
27. Kirak, O. *et al.* Transnuclear mice with pre-defined T cell receptor specificities against *Toxoplasma gondii* obtained via SCNT. *J. Vis. Exp.* **43**, e2168 (2010).
28. Sehwat, S. *et al.* CD8⁺ T cells from mice transnuclear for a TCR that recognizes a single H-2K^b-restricted MHV68 epitope derived from gB-ORF8 help control infection. *Cell Rep.* **1**, 461–471 (2012).
29. Kishigami, S. *et al.* Production of cloned mice by somatic cell nuclear transfer. *Nature Protocols* **1**, 125–138 (2006).
30. Simons, K., Helenius, A., Leonard, K., Sarvas, M. & Gething, M. J. Formation of protein micelles from amphiphilic membrane proteins. *Proc. Natl Acad. Sci. USA* **75**, 5306–5310 (1978).



Extended Data Figure 1 | Flu micelles stain HA-specific B cells. **a**, Schematic for preparation of glycoprotein micelles from HA-SRT^{Alexa 647} virus.

b, Immunoprecipitation of HA-Alexa 647 with anti-Alexa 647 monoclonal antibody. Triton X100-disrupted virions were incubated with 400 µg anti-Alexa 647 overnight and HA-Alexa 647 was then recovered using protein G-Sepharose. Bound proteins were eluted with 0.1 M glycine pH 2.8. W, wash; E, elution. **c**, Typhoon image of the fractions obtained from a linear sucrose

gradient after 20 h centrifugation (107,900g). **d**, Fraction 8 from the sucrose gradient was concentrated and sucrose-depleted by centrifugation over a 30 kDa filter (Amicon UltraCel). The preparation was stained with phosphotungstate and examined by transmission electron microscopy (×150,000 magnification). **e**, Splenocytes from mice infected with A/WSN/33 or control mice were stained with anti-CD19 and HA-Alexa 647 micelles and analysed by cytofluorometry. Plots are representative of 6 mice per group.



Extended Data Figure 2 | FluBI antibody is of the IgG2b subclass. ELISA plates were coated with A/WSN/33-infected MDCK cell lysate and exposed to 1:100 diluted serum from a single C57BL/6 (wt), FluBI, FluBI;*Rag2*^{-/-}, or wild-type mouse infected with A/WSN/33. Plates were washed and probed with

isotype-specific secondary antibodies. Uninfected wild-type mice have flu-reactive antibodies of the IgM subclass. Flu-specific IgE was not detected in any sample. Error bars are s.d. of samples analysed in triplicate.

FluB IgH sequence:

V= IGHV1-50 D=DQ52 J=JH2

```

atgggatggagctgtatcatcctcttcttggtagcaacagctacaggtgtccactctcag
M G W S C I I L F L V A T A T G V H S Q
gtccaactgcagcagcctggggtgagcttgtgaagcctggggcttcagtgaagctgtcc
V Q L Q Q P G A E L V K P G A S V K L S
tgcaaggcttctggctacaccttcaccagctactggatgcagtggtgaaacagaggcct
C K A S G Y T F T S Y W M Q W V K Q R P
ggacagggccttgatggatcgagagattgatccttctgataactattctaaactacaat
G Q G L E W I G E I D P S D N Y S N Y N
caaaagttgaagggaaggccacattgactgtagacacatcctccagcacagcctacatg
Q K L K G K A T L T V D T S S S T A Y M
tacctcagcagcctgacatctgaggactctgcggtctattactgtgcaagatgggagact
Y L S S L T S E D S A V Y Y C A R W E T
Gggtttgactactggggccaaggcaccactctcacagtctcctca
G F D Y W G Q G T T L T V S S

```

FluB IgL sequence:

V= IGKV4-74 J=JK5

```

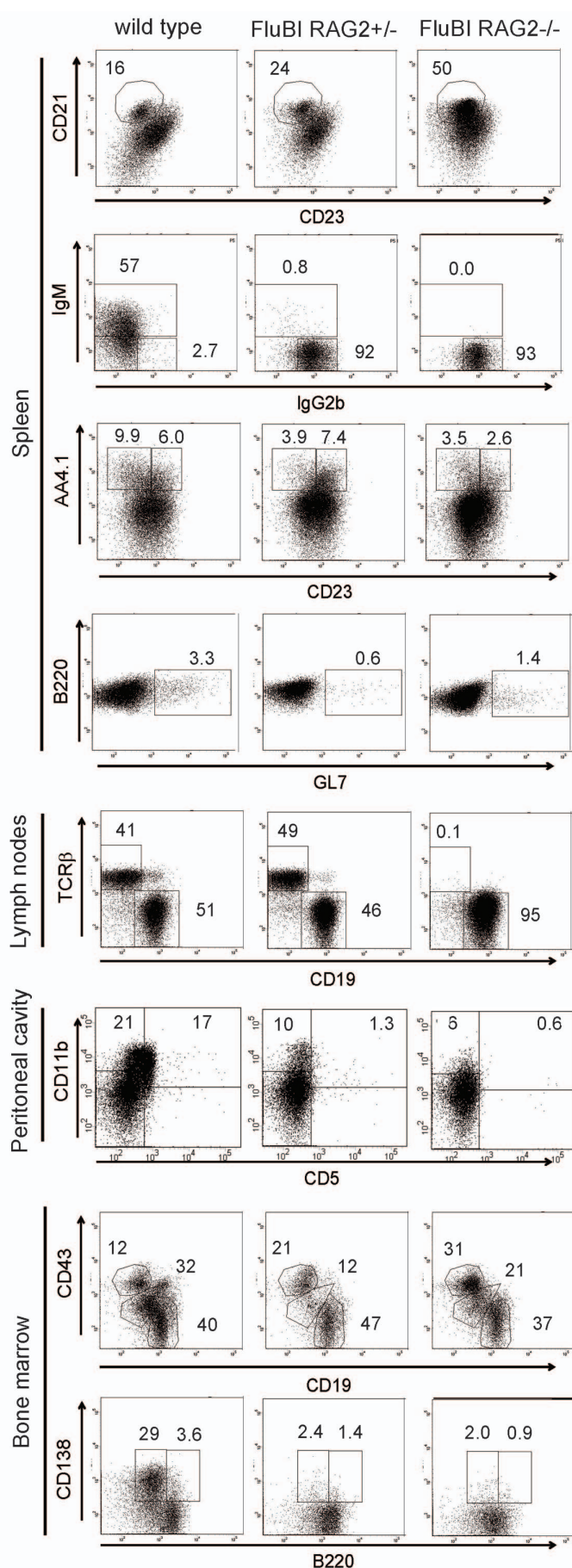
atggattttcaggtgcagattttcagcttctgctaatacagtgccctcagtcataatgtcc
M D F Q V Q I F S F L L I S A S V I M S
agaggacaaattgttctcaccagctctccagcaatcatgtctgcatctctaggggaacgg
R G Q I V L T Q S P A I M S A S L G E R
gtcaccatgacctgcaactgccagctcaagtgtgaagttccaactactgcaactggtaaccag
V T M T C T A S S S V S S T Y L H W Y Q
cagaagccaggatcctcccccaaaagtctggatttatagcacatccaacctggcttctgga
Q K P G S S P K V W I Y S T S N L A S G
gtcccatctcgtcttcagtggcagtggtctgggacctcttactctctcacaaatcagcag
V P S R F S G S G S G T S Y S L T I S S
atggaggctgaagatgctgccacttattactgccaccagtatcatcggtccccacccacg
M E A E D A A T Y Y C H Q Y H R S P P T
ttcgggtgctgggaccaagctggagctgatacgt
F G A G T K L E L I R

```

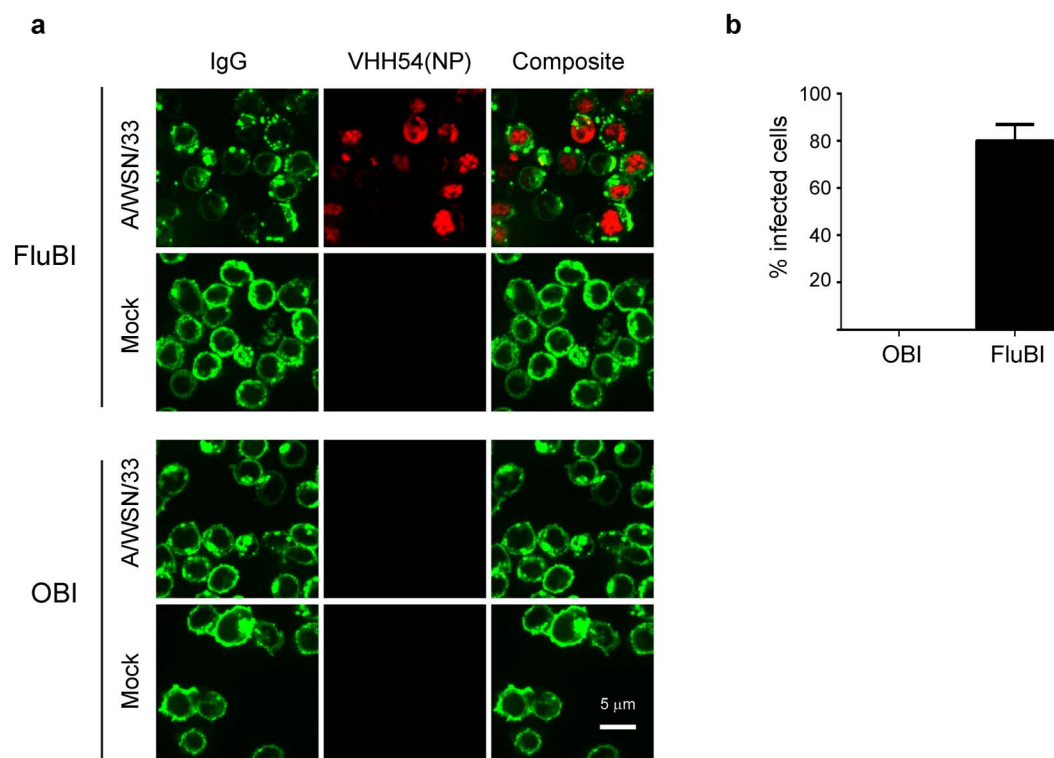
Silent somatic mutations**Somatic mutations resulting in amino acid changes**

Extended Data Figure 3 | Sequence of the VDJ and VJ segments of the FluBI antibody. Genomic DNA was prepared from tails of FluBI mice. The heavy and light chain rearrangements were first identified by amplifying and sequencing of the segments with degenerate primers: for heavy chain: forward 5'-ARGCCTGGGRCTTCAGTGAAG-3' and reverse 5'-AGGCTCTGAGATCCCTAGACAG-3'; for light chain: forward 5'-GGCTGCAGSTTCAGTGGCAGTGGRTCWGGRAC-3' and reverse 5'-ATGCGACGTCAACTGATAATGAGCCCTCTCC-3'. Then the

full sequences of the rearranged heavy and light chain segments were obtained using specific primers: forward 5'-TTACTGAGCACACAGGACCTC-3' and reverse 5'-AGGCTCTGAGATCCCTAGACAG-3'; for light chain: forward 5'-CAGCCCATATTCTCCCATGT-3' and reverse 5'-ATGCGACGTCAACTGATAATGAGCCCTCTCC-3'. Amplified products were agarose gel-purified and sequenced. Sequences were aligned to the NCBI mouse V, D and J genes using IgBlast. Sequences were deposited in GenBank (accession numbers KF419287 and KF419288).

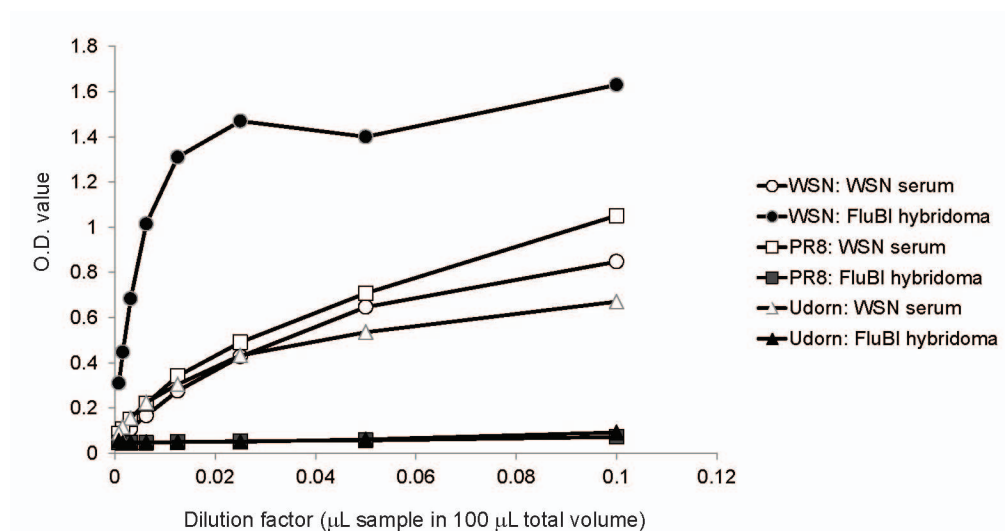


Extended Data Figure 4 | FluBI mice lack B-1a B cells, but show near-normal development of follicular B cells. Cells were isolated from spleen, lymph node (LN, pooled mesenteric and cervical), peritoneal cavity and bone marrow of FluBI, FluBI *Rag2*^{+/-} or C57BL/6 mice. Erythrocytes were lysed and cells were stained with the indicated antibodies and 7-AAD viability dye. LN plots were gated on total live cells. All other populations were gated on CD19⁺ live cells. Numbers indicated the percentage of cells in the indicated gates. B-1a B cells (CD5⁺) are absent and B-1b B cells (CD5⁻CD11b⁺) are reduced in the peritoneal cavity of FluBI and FluBI *Rag2*^{+/-} mice. Plots are representative of 5 mice per group.



Extended Data Figure 5 | FluBI B cells are infected by A/WSN/33. CD40-activated OBI or FluBI B cells were incubated with A/WSN/33 virus at an MOI of 1.0 for 30 min on ice. Cells were then washed and incubated at 37 °C in RPMI (0.2% BSA). At 2 h.p.i., cells were fixed, permeabilized and stained with anti-IgG and TAMRA-conjugated anti-NP (VHH54, derived from alpaca;

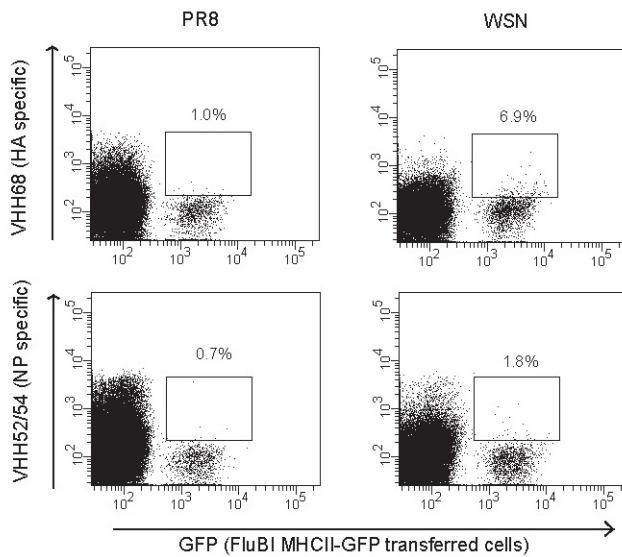
see Extended Data Fig. 9). **a**, Cells were visualized by confocal microscopy. **b**, Cells from **a** were scored as VHH54-positive or -negative. Error bars represent s.d. of positive cells counted per field (3 fields counted; ~200 total cells were counted per group).



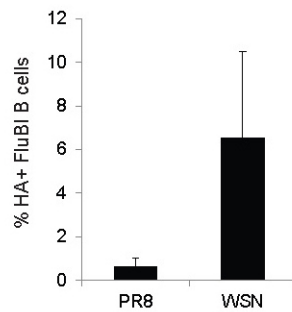
Extended Data Figure 6 | Antibody secreted by FluBI B cells does not cross-react with other strains of influenza virus. ELISA plates were coated with A/WSN/33 (H1N1), A/Udon/307/1972 (H3N2) or A/Puerto Rico/8/1934 (H1N1) overnight at 4°. Plates were then washed, blocked with 10% fetal

bovine serum and exposed to FluBI hybridoma supernatant or WSN-infected serum at the indicated dilutions. Bound antibody was detected using horseradish peroxidase-coupled anti-IgG2b secondary reagent.

a

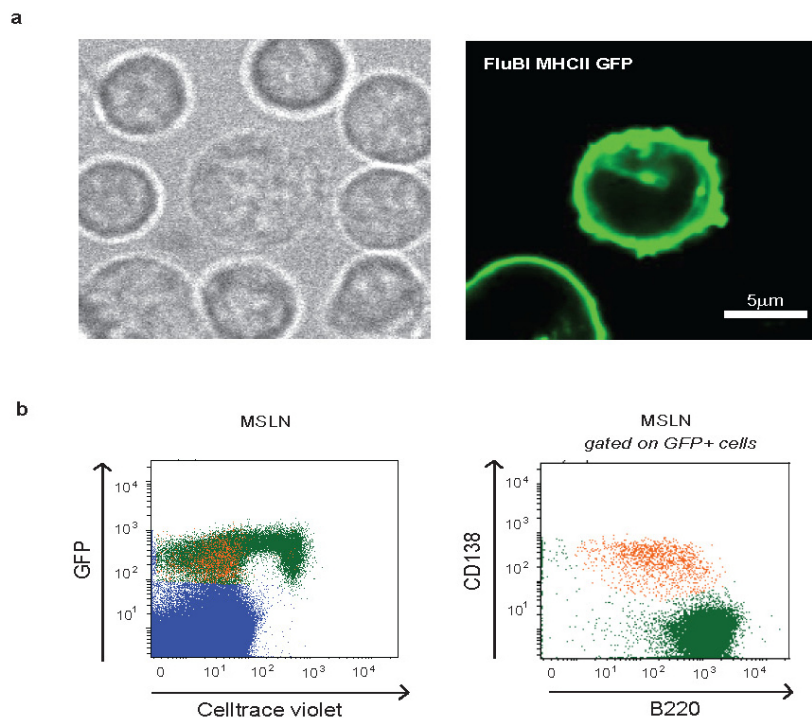


b



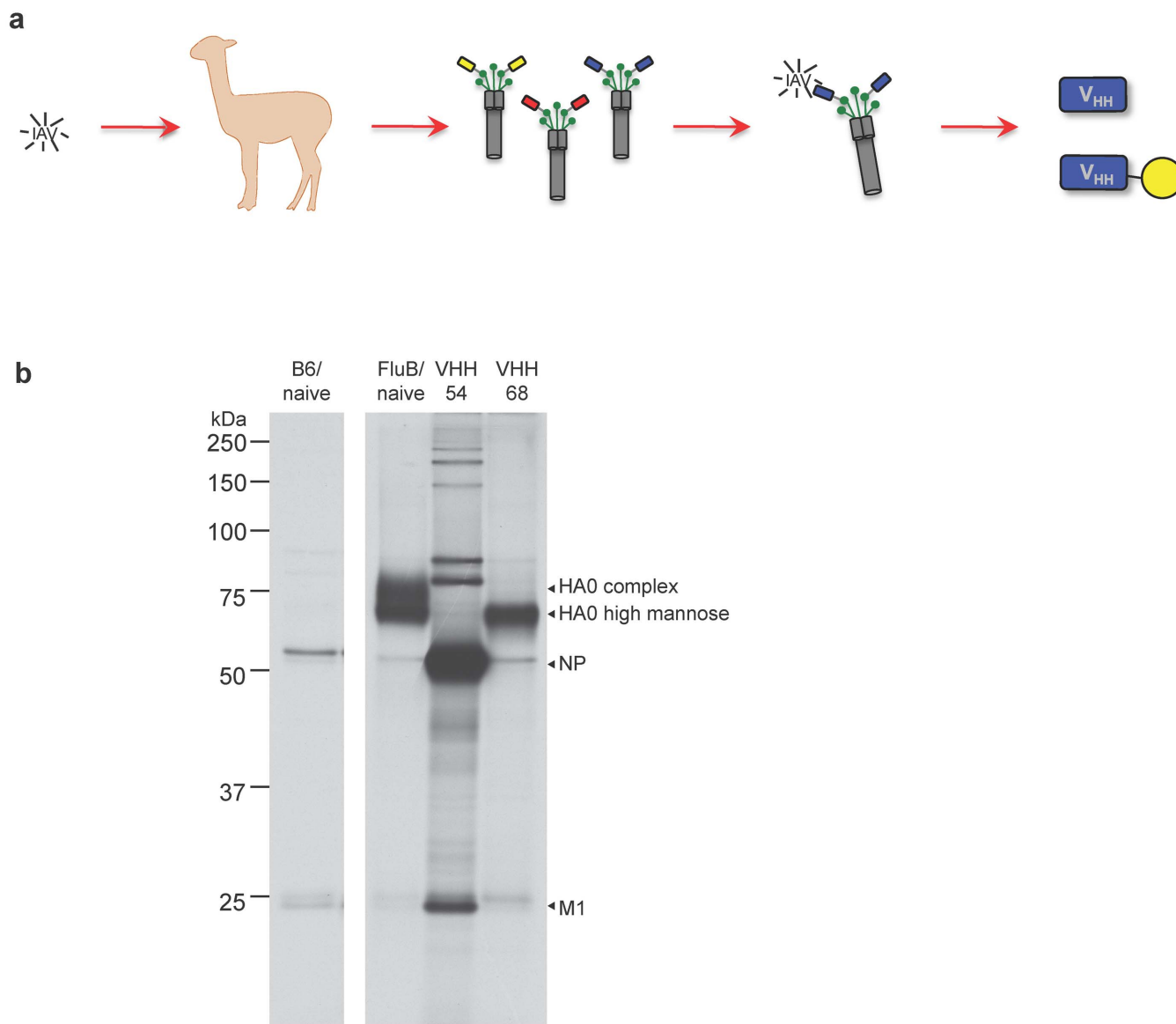
Extended Data Figure 7 | FluBI B cells are not infected with A/Puerto Rico/8/1934 virus *in vivo*. C57BL/6 mice were administered 5×10^6 MHCII-GFP⁺ FluBI B cells 2 h before intranasal infection with 2×10^5 p.f.u. per mouse of either A/WSN/33 (WSN) or A/Puerto Rico/8/1934 (PR8). Mice were euthanized 3 days post-infection, and lung resident cells were stained with anti-CD19 and TAMRA-conjugated VHH68 (anti-HA) or TAMRA-conjugated VHH52/54 (anti-NP). **a**, Representative plots gated on CD19⁺ cells.

b, Quantification of flu-antigen positive cells as shown in **a**. $n = 3$. Error bars are s.d. $p = 0.06$ using two-sided *t*-test.



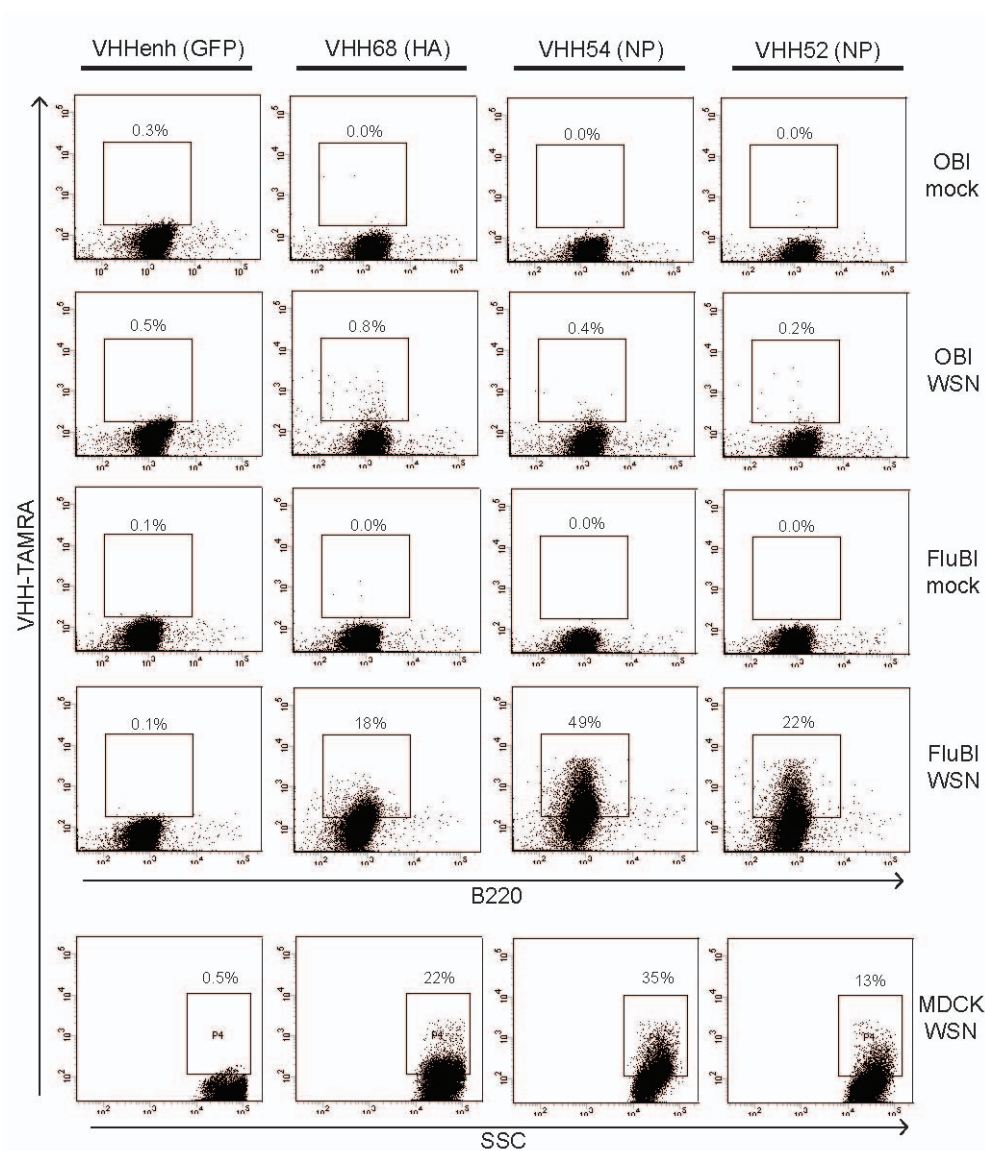
Extended Data Figure 8 | Proliferating FluBI cells in the mediastinal lymph node are plasmablasts. **a**, Mediastinal lymph node cells from day 6 post live infection mice described in Fig. 4 were analysed by confocal microscopy.

GFP⁺ cells displayed a morphology consistent with plasmablasts. **b**, MSLN cells from day 6 post live infection mice described in Figure 4 were analysed by cytofluorometry. Proliferating (violet low) cells were B220^{low} and CD138⁺.



Extended Data Figure 9 | Alpaca-derived VHHS recognize HA and NP from A/WSN/33. **a**, An alpaca was immunized with ethanol-fixed influenza virus. Phage display libraries were constructed from selectively amplified VHH-specific complementary DNA using peripheral blood lymphocytes as starting material, and panned twice against sortase labelled influenza

HA-SRT^{biotin} virus bound to streptavidin coupled beads. VHH sequences obtained from specific binders were expressed with a sortase recognition motif to allow direct conjugation of biotin or fluorophores. **b**, VHH54 and VHH68 conjugated directly to agarose beads were used to precipitate lysates of A/WSN/33 infected, [³⁵S]cysteine/methionine-labelled MDCK cells.



Extended Data Figure 10 | Flu-specific VHs can stain infected FluBI B cells. B cells from OBI or FluBI mice were cultured for 24 h in RPMI containing anti-CD40 ($1 \mu\text{g ml}^{-1}$) before exposure to A/WSN/33. OBI B cells, FluBI B cells and MDCK cells were incubated with A/WSN/33 at an MOI of 1.0 for 30 min on ice, washed once with PBS, and transferred to 37°C in RPMI (0.2% BSA). At

5 h post infection, cells were washed, permeabilized, fixed and stained using TAMRA-conjugated flu-specific VHs ($1 \mu\text{g}$ in $50 \mu\text{l}$). Infected MDCK cells were analysed in parallel as a positive control. Cells were analysed by cytofluorometry using a BD Fortessa.

The nuclear receptor Rev-erb α controls circadian thermogenic plasticity

Zachary Gerhart-Hines^{1,2}, Dan Feng^{1,2*}, Matthew J. Emmett^{1,2*}, Logan J. Everett^{1,2}, Emanuele Loro^{3,4}, Erika R. Briggs^{1,2}, Anne Bugge^{1,2}, Catherine Hou⁵, Christine Ferrara⁶, Patrick Seale^{2,7}, Daniel A. Pryma⁵, Tejvir S. Khurana^{3,4} & Mitchell A. Lazar^{1,2}

Circadian oscillation of body temperature is a basic, evolutionarily conserved feature of mammalian biology¹. In addition, homeostatic pathways allow organisms to protect their core temperatures in response to cold exposure². However, the mechanism responsible for coordinating daily body temperature rhythm and adaptability to environmental challenges is unknown. Here we show that the nuclear receptor Rev-erb α (also known as Nr1d1), a powerful transcriptional repressor, links circadian and thermogenic networks through the regulation of brown adipose tissue (BAT) function. Mice exposed to cold fare considerably better at 05:00 (Zeitgeber time 22) when Rev-erb α is barely expressed than at 17:00 (Zeitgeber time 10) when Rev-erb α is abundant. Deletion of Rev-erb α markedly improves cold tolerance at 17:00, indicating that overcoming Rev-erb α -dependent repression is a fundamental feature of the thermogenic response to cold. Physiological induction of uncoupling protein 1 (Ucp1) by cold temperatures is preceded by rapid downregulation of Rev-erb α in BAT. Rev-erb α represses Ucp1 in a brown-adipose-cell-autonomous manner and BAT Ucp1 levels are high in Rev-erb α -null mice, even at thermoneutrality. Genetic loss of Rev-erb α also abolishes normal rhythms of body temperature and BAT activity. Thus, Rev-erb α acts as a thermogenic focal point required for establishing and maintaining body temperature rhythm in a manner that is adaptable to environmental demands.

The molecular clock is an autoregulatory network of core transcriptional machinery orchestrating behavioural and metabolic programming in the context of a 24-h light–dark cycle^{1,3}. The importance of appropriate synchronization in organismal biology is underscored by the robust correlation between disruption of clock circuitry and development of disease states such as obesity, diabetes mellitus and cancer^{4–6}. Tissue-specific clocks are entrained by environmental stimuli, blood-borne hormonal cues, and direct neuronal input from the suprachiasmatic nucleus located in the hypothalamus to ensure coordinated systemic resonance^{1,7}.

One of the defining metrics of circadian patterning is body temperature⁸, which is highest in animals while awake and lowest while asleep¹. A major site of mammalian thermogenesis is BAT, which is characterized by high glucose uptake, oxidative capacity and mitochondrial uncoupling². Despite a substantial body of literature examining various regulatory aspects of BAT function and body temperature, little is known about the mechanisms controlling circadian thermogenic rhythms and, more importantly, how this patterning influences adaptability to environmental challenges. The circadian transcriptional repressor Rev-erb α has been previously linked to the regulation of glucose and lipid metabolism in tissues such as skeletal muscle, white adipose and liver^{9–15}, but its influence on BAT physiology remains unknown.

We investigated the function of Rev-erb α in controlling temperature rhythms and thermogenic plasticity through integration of circadian and environmental signals. All experiments were performed on C57BL/6 mice and, unless otherwise noted, at murine thermoneutrality (~29–30 °C) to avoid confounding background contributions from the ‘browning’ of white adipose depots or partial stimulation of BAT activity¹⁶. At thermoneutrality, the circadian oscillations of Rev-erb α gene expression (Fig. 1a) and protein levels (Extended Data Fig. 1a) in BAT were similar to other tissues^{11,17}, peaking in the light and being nearly absent in the dark. Rev-erb α ablation altered *Bmal1* (also known as *Arntl*) transcription but did not affect the rhythmicity of Rev-erb β (also known as *Nr1d2*), *Cry1*, *Cry2*, *Per1*, *Per2*, *Per3* or *Clock* (Extended Data Fig. 1b), consistent with the mild circadian phenotype observed previously¹⁷.

To evaluate the role of Rev-erb α in BAT, C57BL/6 wild-type and Rev-erb α knockout mice were subjected to an acute cold challenge from Zeitgeber time (ZT) 4–10 (11:00–17:00) when Rev-erb α levels peak in wild-type animals. In accordance with previous reports that thermoneutrally acclimated C57BL/6 mice fail to thrive during acute cold stresses^{16,18,19}, body temperatures of wild-type animals dropped markedly when shifted from 29 °C to 4 °C (Fig. 1b), and this inability to maintain body temperature was associated with failure to survive the cold exposure (Fig. 1c). By contrast, Rev-erb α knockout mice maintained body temperature and uniformly survived the ZT4–10 cold challenge.

Notably, these studies were all performed during the day, when Rev-erb α peaks in wild-type mice. As Rev-erb α is physiologically nearly absent at night, we next explored whether the circadian expression of Rev-erb α imposed a diurnal variation in cold tolerance. Previous studies of animals exposed to cold at either mid-morning or early afternoon reported modest differences in tolerance, but this effect was believed to be a result of altered vasodilation²⁰. Notably, during the dark period, when Rev-erb α levels are at the nadir of their physiological rhythm, wild-type mice were fully able to protect their body temperature and were phenotypically indistinguishable from Rev-erb α knockout mice in both body temperature regulation (Fig. 1d) and survival (Fig. 1e) following cold challenge. These findings implicate Rev-erb α in establishing a circadian rhythm of cold tolerance through suppression of heat-producing pathways.

The increased cold tolerance of Rev-erb α knockout mice was associated with higher oxygen consumption rates compared to wild-type littermates (Fig. 1f). Food intake (Extended Data Fig. 2a), basal muscle activity and cold-induced shivering (Fig. 1g and Extended Data Fig. 2b) were unchanged between genotypes, indicating that the Rev-erb α -dependent differences in oxidative capacity were probably due to alterations in a BAT-driven, non-shivering thermogenic program. Indeed, BAT isolated from cold-challenged Rev-erb α knockout animals consumed more oxygen than BAT from wild-type mice (Fig. 1h). Moreover,

¹Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ²The Institute for Diabetes, Obesity, and Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ³Department of Physiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴Pennsylvania Muscle Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁵Department of Radiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁶Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. ⁷Department of Cell and Developmental Biology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

*These authors contributed equally to this work.

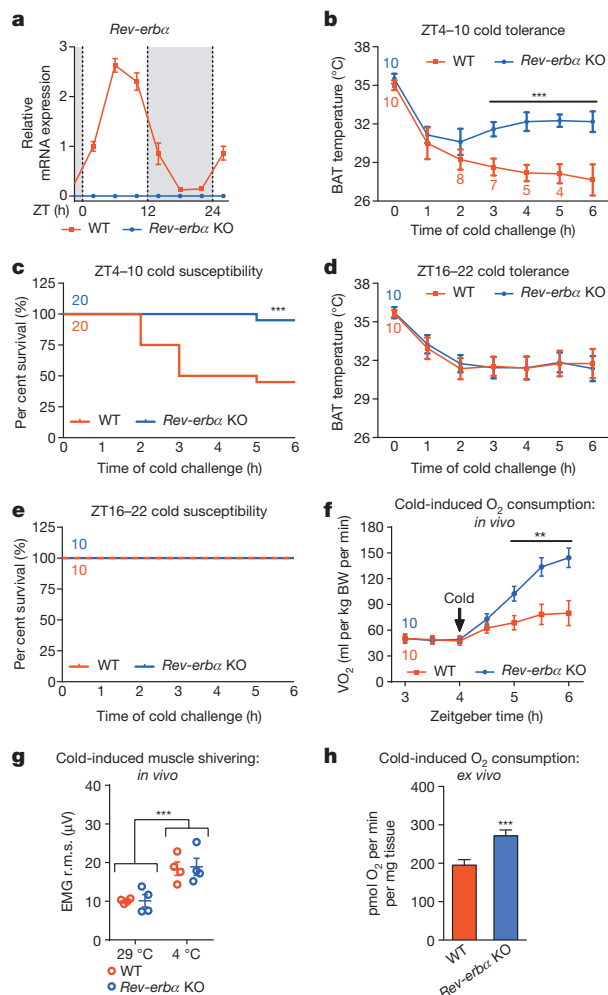


Figure 1 | *Rev-erbα* mediates the circadian patterning of cold tolerance.

a, *Rev-erbα* mRNA ($n = 3$) in BAT of wild-type (WT) and *Rev-erbα* knockout (KO) mice. **b**, **c**, Cold tolerance tests (CTTs) (b) and survival curves (c) for *Rev-erbα* knockout mice and control littermates from ZT4-10 (11:00–17:00). **d**, **e**, CTTs (d) and survival curves (e) from ZT16-22 (23:00–05:00). The numbers of *Rev-erbα* knockout and control mice in the CTT are indicated above or below the first data point, respectively; subsequent designations at data points are made if any animals were removed for having a temperature below 25 °C. **f**, **g**, Oxygen consumption rate ($n = 10$) (f) and Root mean squared (r.m.s.) derivation of electromyogram (EMG) measurements ($n = 4$) (g) of cold-challenged *Rev-erbα* knockout mice and wild-type controls. BW, body weight. **h**, Oxygen consumption rates of BAT isolated from animals exposed to cold for 1 h ($n = 3$). $^{**}P < 0.01$, $^{***}P < 0.001$ as analysed by two-tailed Student's *t*-test, one-way analysis of variance (ANOVA) or Gehan-Breslow-Wilcoxon and log-rank (Mantel-Cox) tests for the survival curves. Data are expressed as mean \pm s.d.

noradrenaline administration induced a larger increase in oxygen consumption in *Rev-erbα* knockout animals than in control littermates (Extended Data Fig. 2c) with no genotypic difference in muscle activity (Extended Data Fig. 2d, e), further suggesting that *Rev-erbα* modulates heat production and cold susceptibility through BAT thermogenic pathways. Despite enhanced BAT metabolic capacity, *Rev-erbα* knockout mice exhibited no statistically significant difference in weight or food intake at room temperature (22 °C) and thermoneutrality compared to wild-type controls (data not shown), probably due to counteracting effects of *Rev-erbα* deletion in other tissues such as increased hepatic lipogenesis⁹ or decreased skeletal muscle oxidative capacity¹⁰.

Given the considerable influence that environmental demands have on BAT-mediated thermogenesis, we investigated whether *Rev-erbα* was subject to control by temperature in BAT. *Rev-erbα* levels normally rise between ZT4 and 10 (11:00 and 17:00) in a circadian manner, but

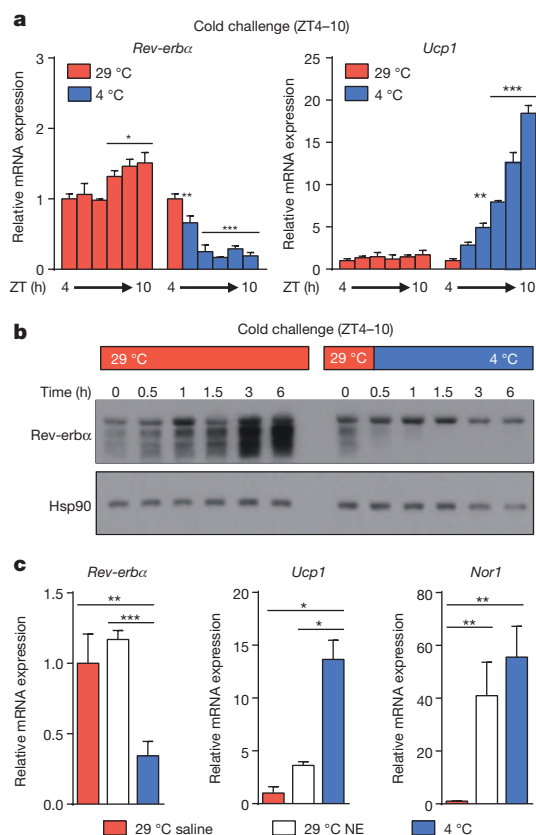


Figure 2 | Cold stress rapidly downregulates *Rev-erbα*. **a**, **b**, BAT mRNA ($n = 3$) (a) and protein levels (b) from wild-type mice following a cold-exposure time course ($n = 3$; each lane of the western blot represents pooled biological duplicates). **c**, BAT mRNA ($n = 3$) following 3 h noradrenaline (NE) administration (1 mg kg⁻¹ i.p.) or cold exposure ($n = 3$). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ as determined by one-way ANOVA with multiple comparisons and a Tukey's post-test. Data are expressed as mean \pm s.d.

cold exposure rapidly attenuated *Rev-erbα* expression (Fig. 2a), whereas closely related nuclear receptor *Rev-erbβ* did not undergo a similar cold-dependent decrease (Extended Data Fig. 3a). Cold-mediated reduction of *Rev-erbα* gene expression occurred in parallel with the induction of *Bmal1*, an established target of *Rev-erbα* repression (Extended Data Fig. 3b), as well as the canonical thermogenic regulators *Ucp1* (Fig. 2a) and peroxisome proliferator-activated receptor gamma coactivator 1 alpha (*Pgc-1α*, also known as *Ppargc1a*)²¹ (Extended Data Fig. 3c). *Rev-erbα* expression was attenuated following both moderate (29 °C to 20 °C) and acute (29 °C to 4 °C) cold stresses (Extended Data Fig. 3d). Similarly, *Rev-erbα* protein levels plummeted when mice were shifted to 4 °C (Fig. 2b). Classically, regulation of brown adipose thermogenesis has been attributed predominantly to sympathetic release of noradrenaline and subsequent activation of adrenergic signalling cascades². We therefore considered whether the cold-induced decrease in *Rev-erbα* levels was related to the adrenergic pathway. However, whereas the highly cyclic-AMP-sensitive nuclear receptor *Nor1* (also known as *Nr4a3*) (ref. 22) was induced comparably by noradrenaline and cold (Fig. 2c), noradrenaline administration did not mimic the effect of cold exposure on expression of *Rev-erbα* gene (Fig. 2c) or protein (Extended Data Fig. 3e). This is consistent with reports that pan-sympathomimetic stimulation does not fully recapitulate cold-mediated BAT activation in humans^{23,24}, and suggests that the role of *Rev-erbα* in thermogenic regulation is independent of sympathetic stimulation.

The rapidity with which *Rev-erbα* was reduced in the cold and its inverse relationship with *Ucp1* expression suggested that *Rev-erbα* might elicit thermogenic regulation through active repression of the *Ucp1* gene. Indeed, at thermoneutrality *Ucp1* mRNA (Fig. 3a) and protein levels (Fig. 3b) were increased in the BAT of *Rev-erbα* knockout

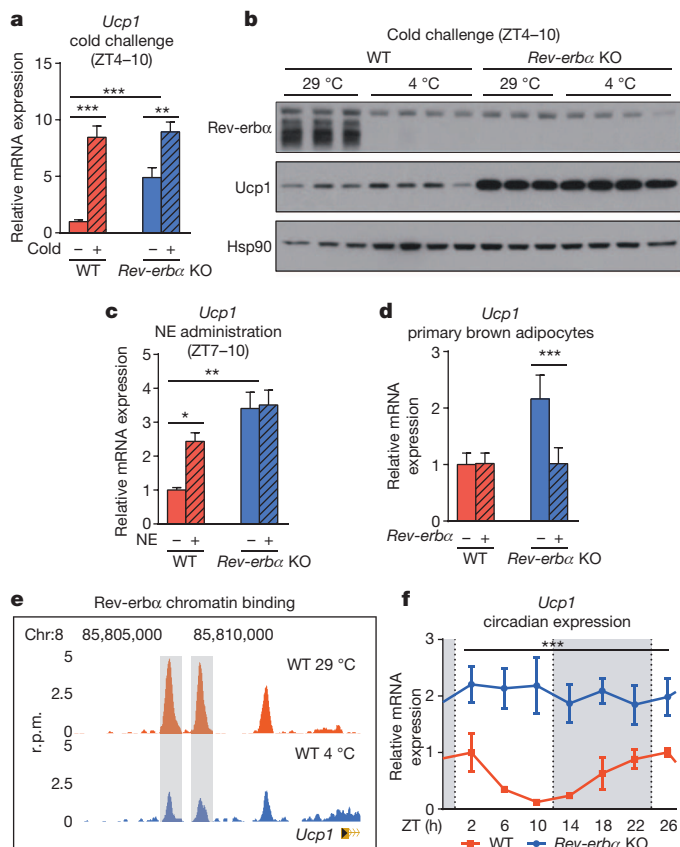


Figure 3 | Rev-erb α represses thermogenic programming. **a, b**, BAT mRNA ($n = 6$) (**a**) and protein (**b**) from wild-type and *Rev-erb α* knockout mice acutely exposed to cold for 6 h from ZT4–10. **c**, BAT mRNA following 3 h of noradrenaline administration from ZT7–10 (1 mg kg⁻¹ i.p.) ($n = 3$). **d**, *Ucp1* mRNA levels in preadipocytes isolated from *Rev-erb α* knockout mice and wild-type littermates in which either *Rev-erb α* or vector control has been ectopically expressed ($n = 4$). **e**, *Rev-erb α* occupancy at the *Ucp1* proximal promoter. *Rev-erb α* -specific peaks are shaded. **f**, *Ucp1* gene expression in BAT over a 24-h period ($n = 3$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ as determined by two-tailed Student's *t*-test or one-way ANOVA with multiple comparisons and a Tukey's post-test. Data are expressed as mean \pm s.d.

mice, consistent with the more pronounced metabolic response of these mice to cold exposure or noradrenaline administration. BAT *Ucp1* in *Rev-erb α* knockout mice was only modestly further increased upon cold challenge compared to wild-type animals (Fig. 3a, b), suggesting that *Rev-erb α* downregulation is an integral component of the physiological *Ucp1* induction following cold exposure. Consistent with recent work on the temporal correlation between *Ucp1* mRNA and protein levels, we did not observe cold-mediated changes in *Ucp1* protein in the acute time frame in which we performed our cold challenges²⁵. Increases in *Ucp1* were not seen in white adipose depots or skeletal muscle (data not shown), signifying a BAT-specific phenomenon. *Bmal1* mRNA and protein followed a similar pattern to *Ucp1* (Extended Data Fig. 4a, b), whereas *Pgc-1 α* levels were unchanged between control and *Rev-erb α* knockout animals at thermoneutrality, and were comparably cold-induced, suggesting *Rev-erb α* independence (Extended Data Fig. 4a, b). Nevertheless, *Rev-erb α* controlled the expression of *Ucp1*, which is critical for non-shivering heat production in BAT^{2,19}. Underscoring this point, *Ucp1* mRNA levels were basally higher in *Rev-erb α* knockout mice given only saline than those of noradrenaline-treated wild-type animals and were not increased further when noradrenaline was administered to the *Rev-erb α* knockouts (Fig. 3c).

Ucp1 levels were increased in primary brown adipocytes lacking *Rev-erb α* and ectopic expression of *Rev-erb α* restored *Ucp1* mRNA to wild-type levels, whereas overexpression of *Rev-erb α* in wild-type adipocytes caused

no further effect (Fig. 3d), illustrating that *Rev-erb α* represses *Ucp1* in a BAT-cell-autonomous manner. Consistent with these findings, *Rev-erb α* binding was detected at the *Ucp1* gene locus, and this binding decreased after cold challenge (Fig. 3e). *Ucp1* displayed a rhythmic expression profile anti-phase to *Rev-erb α* in primary brown adipocytes cultured *ex vivo* and synchronized by serum shock (Extended Data Fig. 4c). This *Ucp1* circadian rhythmicity was completely abolished in *Rev-erb α* knockout animals (Fig. 3f). These data establish *Rev-erb α* as a direct, negative regulator of thermogenic transcriptional programs.

The ability of *Rev-erb α* to repress BAT heat production and impose a circadian pattern of cold tolerance prompted us to investigate whether *Rev-erb α* influenced body temperature rhythm. *Rev-erb α* ablation considerably altered body temperature oscillation, both of the core (Fig. 4a) and of the interscapular region (BAT) (Fig. 4b). Higher body temperature was maintained by *Rev-erb α* knockout animals throughout the light phase, indicating that *Rev-erb α* was required for daily depressions in thermogenic rhythmicity. Indeed, thermographic surface measurements showed that *Rev-erb α* knockout mice were warmer than wild-type mice from ZT4–10 (11:00–17:00) but not ZT16–22 (23:00–05:00) (Fig. 4c and Extended Data Fig. 5a). Comparison between colonic and interscapular temperatures implicated BAT as the primary source of the genotypic variation (Extended Data Fig. 5a). We note that previous studies of thermoregulation in mice lacking *Rev-erb α* were performed at room temperature, which could confound the assessment of the role of *Rev-erb α* in BAT thermogenesis^{13,16}.

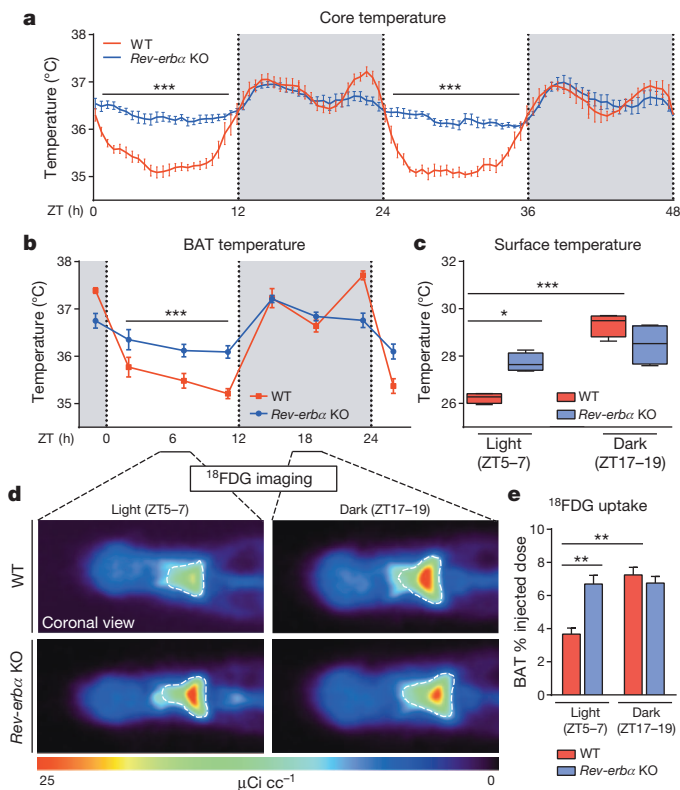


Figure 4 | Rev-erb α orchestrates daily rhythms of body temperature and BAT activity. **a, b**, Core ($n = 6$) (**a**) and BAT ($n = 10$) (**b**) temperatures measured from subcutaneously implanted thermometers. **c, d**, Quantified thermographic measurements of surface temperature ($n = 5$) (**c**) and ¹⁸F-FDG imaging ($n = 4$) (**d**) of *Rev-erb α* knockout mice and wild-type littermates during the light and dark phases. Representative coronal planes are shown for each group. **e**, Per cent injected dose of ¹⁸F-FDG in the BAT of animals from the study in **d**. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ as determined by two-tailed Student's *t*-test or one-way ANOVA with multiple comparisons and a Tukey's post-test. Data in **a** are expressed as rolling averages (± 2 time points) \pm s.e.m.; data in **b** are expressed as mean \pm s.e.m.; data in **c** are expressed as a maximum to minimum box-and-whiskers plot; data in **e** are expressed as a mean \pm s.d.

To address the effect of *Rev-erb α* on the circadian control of BAT function, we measured glucose uptake using 18-fluorodeoxyglucose positron emission tomography (^{18}F FDG-PET)^{26–28}. Notably, the diurnal oscillation of BAT glucose uptake²⁹ was abolished by deletion of *Rev-erb α* (Fig. 4d and Extended Data Fig. 5c). Glucose uptake was higher in *Rev-erb α* knockout mice than control littermates during the day and did not increase at night as in wild-type animals (Fig. 4e). These results indicate that *Rev-erb α* is required for the circadian rhythm of body temperature and BAT activity (Extended Data Fig. 6).

Daily oscillation in body temperature is one of the most basic and defining characteristics of mammalian circadian biology⁶. The present findings suggest a mechanism whereby circadian and cold-regulated networks converge on *Rev-erb α* in BAT to establish and maintain thermogenic rhythmicity while affording the organism an adaptability to rapidly respond to external temperature stresses. *Rev-erb α* acts as a focal point, integrating the continuity of circadian rhythms with the variability of environmental challenges. *Rev-erb α* alone is sufficient to modulate brown adipose function, which is in contrast to the redundancy found between both nuclear receptors *Rev-erb α* and *Rev-erb β* in controlling hepatic physiology^{11,12}. The fact that *Rev-erb β* is not subject to similar cold-dependent regulation ensures that temperature stresses can target appropriate programs without detriment to the BAT core clock machinery.

The function of BAT as a professional heat-producing tissue probably evolved to permit eutherian mammals to survive exposure to an array of environmental demands³⁰. However, from an evolutionary standpoint, constitutive, Ucp1-mediated dissipation of the mitochondrial proton gradient would be wasteful and unfavourable when resources are scarce and increased heat production is unnecessary. Our data are consistent with a model of *Rev-erb α* -controlled BAT thermogenesis that provides an energetic checks-and-balances system. Circadian rhythm of *Rev-erb α* imposes an oscillation in brown adipose activity, increasing body temperature when mammals are awake and potentially exposed to harsh environmental conditions and depressing thermogenesis during sleep when mammals are typically in protective shelter and require little facultative heat production. In the event that the animal is confronted by a sudden temperature challenge while sleeping, rapid reduction in *Rev-erb α* would facilitate appropriate induction of thermogenic programs and organismal survival.

METHODS SUMMARY

Mice were housed on a 12:12-h light–dark cycle (lights on at 07:00, off at 19:00). Gene expression, protein analysis and temperature measurements were carried out on 12–16-week-old male *Rev-erb α* knockout mice and wild-type littermates. Cold exposure experiments were performed in climate-controlled rodent incubators set to 29 °C and 4 °C. Oxygen consumption rates were measured using Comprehensive Laboratory Animal Monitoring System (CLAMS) metabolic cages contained with temperature-controlled rodent incubators. Core and brown adipose temperature measurements were obtained using surgically implanted dataloggers for core (SubCue Dataloggers) and telemetric transmitters for BAT (IPTT 300 transponders, Bio Medic Data Systems) following pentobarbital anaesthetization. Colonic and interscapular surface measurements were obtained using YSI Precision Thermometers with rectal or banjo probe attachments, respectively. Thermography was performed by the Penn Mouse Phenotyping, Physiology, and Metabolism (MPPM) core using a FLIR SC620 infrared camera. ^{18}F FDG imaging was performed in the University of Pennsylvania Small Animal Imaging Facility (SAIF). EMG recordings were performed as described previously¹⁹. Chromatin immunoprecipitation of *Rev-erb α* was performed using the Cell Signaling Technology antibody (no. 2124) as described previously¹¹. Data are presented as means \pm s.d. unless otherwise noted.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 March; accepted 9 September 2013.

Published online 27 October 2013.

1. Bass, J. Circadian topology of metabolism. *Nature* **491**, 348–356 (2012).
2. Cannon, B. & Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiol. Rev.* **84**, 277–359 (2004).

3. Takahashi, J. S., Hong, H.-K., Ko, C. H. & McDearmon, E. L. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nature Rev. Genet.* **9**, 764–775 (2008).
4. Sahar, S. & Sassone-Corsi, P. Metabolism and cancer: the circadian clock connection. *Nature Rev. Cancer* **9**, 886–896 (2009).
5. Asher, G. & Schibler, U. Crosstalk between components of circadian and metabolic cycles in mammals. *Cell Metab.* **13**, 125–137 (2011).
6. Bass, J. & Takahashi, J. S. Circadian integration of metabolism and energetics. *Science* **330**, 1349–1354 (2010).
7. Feng, D. & Lazar, M. A. Clocks, metabolism, and the epigenome. *Mol. Cell* **47**, 158–167 (2012).
8. Buhr, E. D., Yoo, S.-H. & Takahashi, J. S. Temperature as a universal resetting cue for mammalian circadian oscillators. *Science* **330**, 379–385 (2010).
9. Feng, D. *et al.* A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* **331**, 1315–1319 (2011).
10. Woldt, E. *et al.* *Rev-erb α* modulates skeletal muscle oxidative capacity by regulating mitochondrial biogenesis and autophagy. *Nature Med.* (2013).
11. Bugge, A. *et al.* *Rev-erb α* and *Rev-erb β* coordinately protect the circadian clock and normal metabolic function. *Genes Dev.* **26**, 657–667 (2012).
12. Cho, H. *et al.* Regulation of circadian behaviour and metabolism by *Rev-erb α* and *Rev-erb β* . *Nature* **485**, 123–127 (2012).
13. Delezie, J. *et al.* The nuclear receptor *Rev-erb α* is required for the daily balance of carbohydrate and lipid metabolism. *FASEB J.* **26**, 3321–3335 (2012).
14. Le Martelot, G. *et al.* *Rev-erb α* participates in circadian SREBP signaling and bile acid homeostasis. *PLoS Biol.* **7**, e1000181 (2009).
15. Solt, L. A. *et al.* Regulation of circadian behaviour and metabolism by synthetic *Rev-erb* agonists. *Nature* **485**, 62–68 (2012).
16. Cannon, B. & Nedergaard, J. Nonshivering thermogenesis and its adequate measurement in metabolic studies. *J. Exp. Biol.* **214**, 242–253 (2011).
17. Preitner, N. *et al.* The orphan nuclear receptor *Rev-erb α* controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**, 251–260 (2002).
18. Lim, S. *et al.* Cold-induced activation of brown adipose tissue and adipose angiogenesis in mice. *Nature Protocols* **7**, 606–615 (2012).
19. Golozoubova, V. *et al.* Only UCP1 can mediate adaptive nonshivering thermogenesis in the cold. *FASEB J.* **15**, 2048–2050 (2001).
20. Talan, M. I., Tatelman, H. M. & Engel, B. T. Cold tolerance and metabolic heat production in male C57BL/6J mice at different times of day. *Physiol. Behav.* **50**, 613–616 (1991).
21. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
22. Pearen, M. A. *et al.* The orphan nuclear receptor, NOR-1, is a target of β -adrenergic signaling in skeletal muscle. *Endocrinology* **147**, 5217–5227 (2006).
23. Cyppess, A. M. *et al.* Cold but not sympathomimetics activates human brown adipose tissue *in vivo*. *Proc. Natl Acad. Sci. USA* **109**, 10001–10005 (2012).
24. Vosselman, M. J. *et al.* Systemic β -adrenergic stimulation of thermogenesis is not accompanied by brown adipose tissue activity in humans. *Diabetes* **61**, 3106–3113 (2012).
25. Nedergaard, J. & Cannon, B. UCP1 mRNA does not produce heat. *Biochim. Biophys. Acta* **1831**, 943–949 (2013).
26. Cyppess, A. M. *et al.* Identification and importance of brown adipose tissue in adult humans. *N. Engl. J. Med.* **360**, 1509–1517 (2009).
27. Van Marken Lichtenbelt, W. D. *et al.* Cold-activated brown adipose tissue in healthy men. *N. Engl. J. Med.* **360**, 1500–1508 (2009).
28. Virtanen, K. A. *et al.* Functional brown adipose tissue in healthy adults. *N. Engl. J. Med.* **360**, 1518–1525 (2009).
29. Van der Veen, D. R., Shao, J., Chapman, S., Leevy, W. M. & Duffield, G. E. A diurnal rhythm in glucose uptake in brown adipose tissue revealed by *in vivo* PET-FDG imaging. *Obesity* **20**, 1527–1529 (2012).
30. Saito, S., Saito, C. T. & Shingai, R. Adaptive evolution of the uncoupling protein 1 gene contributed to the acquisition of novel nonshivering thermogenesis in ancestral eutherian mammals. *Gene* **408**, 37–44 (2008).

Acknowledgements We thank the Functional Genomics Core (J. Schug) and the Mouse Phenotyping, Physiology, and Metabolism Core (R. Ahima and R. Dhir) of the Penn Diabetes Research Center (NIH P30 DK19525). We also thank the Small Animal Imaging Facility of the Perelman School of Medicine at the University of Pennsylvania (E. Blankemeyer). This work was supported by NIH grants R01 DK45586 (M.A.L.) and F-32 DK095563 (Z.G.-H.) and the JPB Foundation. A.B. was funded by the Novo Nordisk STAR postdoctoral program.

Author Contributions D.F., M.J.E., L.J.E., E.R.B., A.B. and C.F. performed key experiments/data analysis and read the manuscript. P.S. provided advice and read the manuscript. E.L. and T.S.K. designed, performed and analysed EMG studies and read the manuscript. C.H. and D.A.P. designed, performed and analysed ^{18}F FDG scans and read the manuscript. Z.G.H. performed many of the experiments, and Z.G.H. and M.A.L. conceived the project, designed experiments, analysed all results and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.L. (lazar@mail.med.upenn.edu).

METHODS

Animal studies. All animal studies were performed with an approved protocol from the University of Pennsylvania Perelman School of Medicine Institutional Animal Care and Use Committee. The *Rev-erb α* knockout mice were obtained from B. Vennström and backcrossed seven or more generations with C57BL/6 mice. Mice were housed on a 12:12-h light–dark cycle (lights on at 07:00, lights off at 19:00). Gene expression, protein analysis and temperature measurements were carried out on 12–16-week-old male *Rev-erb α* knockout mice and wild-type littermates. Cold-exposure experiments were performed in climate-controlled rodent incubators set to 29 °C and 4 °C. All wild-type and *Rev-erb α* knockout mice used in the studies were first placed in individual cages with access to food and water and allowed to acclimate to 29 °C for 2 weeks before cold challenge. For noradrenaline administration experiments, thermoneutrally acclimated wild-type and *Rev-erb α* knockout mice were given 1 mg kg^{−1} L-(−)-noradrenaline-bitartrate salt monohydrate (Sigma). Mice were injected subcutaneously for noradrenaline-induced oxygen-consumption assays but intraperitoneally for all other procedures.

Whole-animal oxygen-consumption rate. Oxygen-consumption rates were measured using Comprehensive Laboratory Animal Monitoring System (CLAMS) metabolic cages contained within temperature-controlled rodent incubators. Cold-induced oxygen consumption rates were assessed on singly-housed, unanaesthetized wild-type and *Rev-erb α* knockout mice. Temperature of the housing unit was transitioned from 29 °C to 4 °C over the course of 20–30 min, and mice were then cold-challenged for an additional 2 h. Noradrenaline-induced oxygen consumption rates were assessed as described previously¹⁶. In brief, mice were anaesthetized with 75 mg kg^{−1} pentobarbital intraperitoneally and placed in a CLAMS unit set to 33 °C to maintain body temperature. One mg kg^{−1} noradrenaline was administered subcutaneously once a baseline oxygen consumption rate had been obtained (approximately 20 min after pentobarbital injection). Noradrenaline-induced oxygen consumption was then measured until rates had peaked and started declining (approximately 90 min after noradrenaline administration).

Temperature measurements. Core and brown adipose temperature measurements were obtained using surgically implanted dataloggers for core (SubCue Dataloggers) and telemetric transmitters for BAT (IPTT 300 transponders, Bio Medic Data systems) following pentobarbital anesthetization. Mice were maintained at 29 °C and monitored daily and surgical sites were treated with bacitracin to prevent discomfort. Following a week of convalescence, temperature measurements were recorded. Colonic and interscapular surface measurements were obtained using YSI Precision Thermometers with rectal or banjo probe attachments, respectively.

Immunoblotting. BAT samples were homogenized in tissue lysis buffer (137 mM NaCl, 0.1% SDS, 0.5% sodium-deoxycholate, 1% NP-40, 20 mM NaF and 20 mM β -glycerophosphate in 1× PBS, pH 7.4, supplemented with Complete protease inhibitor (Roche)) using a TissueLyser (Qiagen) for 1.5 min at a frequency of 20 s^{−1} followed by sonication using a Bioruptor (Diagenode) for 30 s on the 'high' setting. SDS–PAGE was performed using 50 mg of protein loaded onto a 10% Tris-glycine gel (Invitrogen), followed by transfer to a polyvinylidene difluoride membrane (Invitrogen). After antibody incubation, blots were developed using the SuperSignal West Dura chemiluminescence kit from Pierce.

Cell culture. Preadipocytes were collected from BAT depots of pups that were between postnatal days 1–3. Depots were minced finely using spring scissors (Roboz) in DMEM/F-12 GlutaMax (Invitrogen) before addition of 1.5 U ml^{−1} collagenase D (Roche) and 2.4 U ml^{−1} Dispase II (Roche) and incubation in a 37 °C shaking water bath for 45 min. Cells were purified through 100- μ m filters (Millipore), pelleted and re-suspended in Growth media (DMEM/F-12 GlutaMax supplemented with 10% FBS (Tissue Culture Biologicals), HEPES, pH 7.2 (Invitrogen) and penicillin/streptomycin (Invitrogen)). Adipocyte differentiation was induced upon confluence with induction media (growth media supplemented with 500 nM dexamethasone, 125 nM indomethacin, 0.5 mM IBMX, 1 nM rosiglitazone, 1 nM T3 and 20 nM insulin) for 36 h. After induction, cells were cultured in maintenance media (growth media supplemented with 1 nM T3 and 20 nM insulin). Serum synchronization was performed by incubating differentiated adipocytes with DMEM/F-12 GlutaMax containing 50% horse serum for 2 h. After two washes in PBS, cells were placed in DMEM/F-12 GlutaMax containing 0.5% FBS, 1 nM T3 and 20 nM insulin, and total RNA was collected at the indicated time points. For ectopic *Rev-erb α* expression, primary adipocytes were electroporated 36 h after removing induction media using an Amaxa Cell Line Nucleofector Kit L (Lonza) according to manufacturer's instructions and collected 48 h later.

Thermographic imaging. Thermography was performed by the Penn Mouse Phenotyping, Physiology, and Metabolism (MPPM) core during the light and dark phases using a FLIR SC620 infrared camera on wild-type and *Rev-erb α* knockout mice acclimated at thermoneutrality for 2 weeks. No anaesthesia was used in order to avoid confounding effects on body temperature.

¹⁸FDG imaging. ¹⁸FDG imaging was performed in the University of Pennsylvania Small Animal Imaging Facility (SAIF). Doses of saline containing 300 μ Ci ¹⁸FDG were administered through the lateral tail vein under constant isoflurane anaesthesia (1–2%, 1 L O₂ min^{−1}). Mice were scanned on a Philips Mosaic HP 1 h after injection. Per cent injected dose was calculated by assessing the ratio of radioactive counts in the region of interest for brown adipose to the total counts for the animal using Amide medical imaging software.

BAT oxygen-consumption rate. Mice were housed at thermoneutrality for 1 week and subjected to a 1-h cold challenge (4 °C) starting at 13:00. The interscapular BAT depot of each mouse was collected and divided into 11 pieces, weighing between 1.5 and 2 mg, and washed three times in Seahorse XF assay media supplemented with 25 mM glucose and 1 mM sodium pyruvate and adjusted to pH 7.4. Subsequently, the BAT pieces were placed individually in the centre of a well of a Seahorse XF24 islet capture microplate and held in place by overlaying a capture screen followed by addition of 675 μ l of the supplemented Seahorse XF assay media. The oxygen-consumption rate of each well was measured three times for 2 min after 3 min of mixing and a 2-min wait on the Seahorse XF24 analyser (Seahorse Bioscience). The results from the 11 wells of each genotype were averaged and normalized to total mg of tissue.

EMG. EMG recordings were made essentially as described previously¹⁹. Three 29-gauge needle electrodes (two recording electrodes 4 mm apart and 3 mm deep, and one reference electrode placed distally) were fixed transcutaneously for acquiring the EMG signal from the scapular muscles. For optimal stability, recording electrodes were placed into 4-mm diameter plastic tubes (1 ml serological pipettes) and juxtaposed using polyolefin tubing. The entire electrode set was introduced into the scapular region of prone mice using a micromanipulator (WPI). The EMG signal was processed (low-pass filter, 3 kHz; high-pass filter, 10 Hz; notch filter, 60 Hz) and amplified 1,000× with a P55 differential amplifier (Grass Instruments). Data were A/D converted and recorded with a PowerLab 8SP at a sampling frequency of 10 kHz (ADInstruments). The signal was acquired and r.m.s. of the EMG signal was calculated with LabChart 7 (ADInstruments).

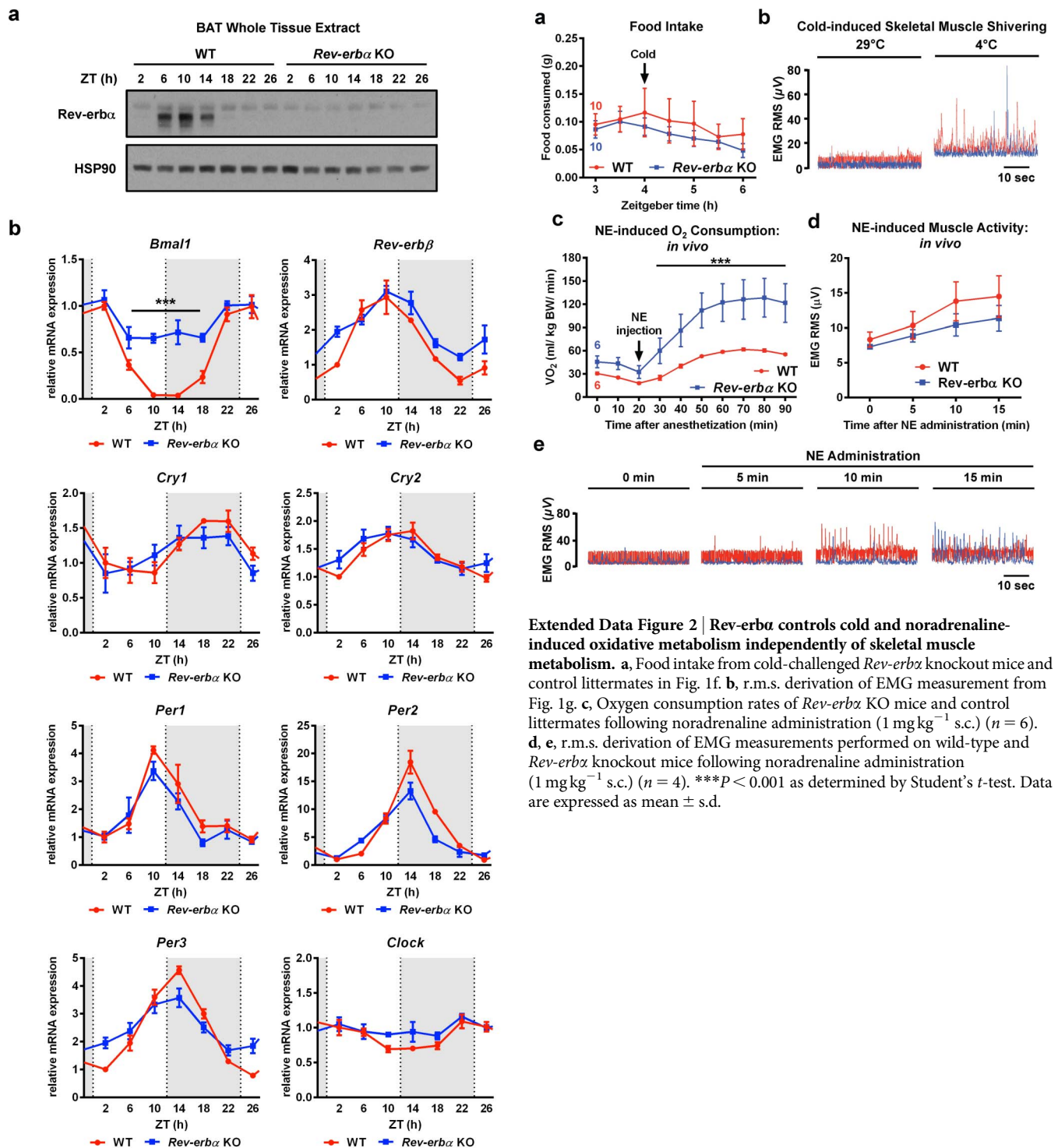
For cold-induced shivering, mice were exposed to 4 °C for 1 h, quickly anaesthetized with isoflurane and placed on a temperature-controlled pad maintained at 15 °C. EMG signals were recorded for 15 min and the data collected between minutes 2 and 7 were used for the analyses. Mice were allowed to recover for 1 day and then subjected to EMG measurement at thermoneutrality, maintaining the temperature-controlled pad at 33 °C.

For recording norepinephrine-induced EMGs, mice were anaesthetized with an intraperitoneal injection of 75 mg kg^{−1} pentobarbital. The temperature-controlled pad was maintained at 33 °C. After obtaining 5 min of basal EMG recordings, 1 mg kg^{−1} noradrenaline was injected subcutaneously on the back of the mouse and the recording continued for 20 min. All r.m.s. calculations were made from 2 min of data collected before noradrenaline administration as well as 5, 10 and 15 min after noradrenaline administration.

ChIP. Murine BAT was collected immediately after euthanasia. It was quickly minced and cross-linked in 1% formaldehyde for 20 min, followed by quenching with 1/20 volume of 2.5 M glycine solution and two washes with ice-cold PBS. Chromatin fragmentation was performed by sonication in ChIP SDS lysis buffer (50 mM HEPES, 1% SDS, 10 mM EDTA, pH 7.5) using probe sonication. Proteins were immunoprecipitated in ChIP dilution buffer (50 mM HEPES, 155 mM NaCl, 1.1% Triton X-100, 0.11% sodium-deoxycholate, complete protease inhibitor tablet, pH 7.5). Crosslinking was reversed overnight at 65 °C in elution buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS, pH 8.0) and DNA was isolated using phenol/chloroform/isoamyl alcohol. Precipitated DNA was analysed by quantitative PCR. ChIP experiments were performed independently on BAT samples from three mice collected at 5 pm with or without a 6-h cold challenge as described previously¹¹. ChIP of *Rev-erb α* was performed using the Cell Signaling Technology antibody. Deep sequencing was carried out by the Functional Genomics Core (J. Schug and K. Kaestner) of the Penn Institute for Diabetes, Obesity, and Metabolism using the Illumina Genome Analyzer Iix and Illumina HiSeq 2000 and sequences were obtained using the Solexa Analysis Pipeline.

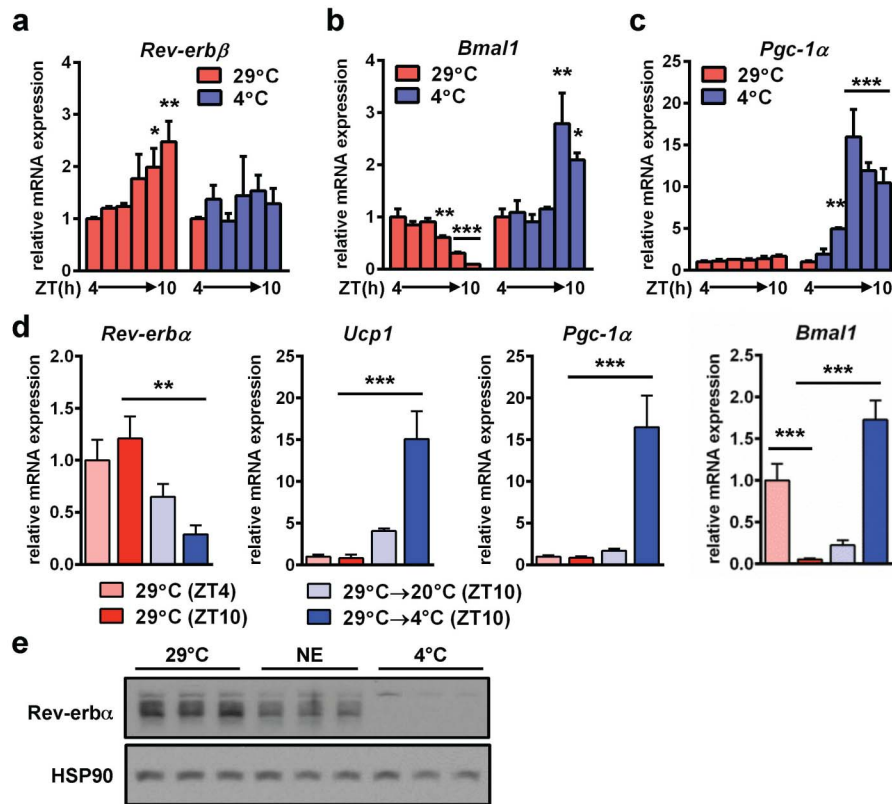
RNA. Total RNA was isolated from BAT tissue by Trizol (Invitrogen) extraction and 1.5 μ g of total RNA was used for complementary DNA synthesis using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems). Relative mRNA levels were determined using quantitative PCR and normalization to housekeeping gene *36B4*. Primer sequences are available upon request.

Statistics. Data are presented as means \pm s.d. unless otherwise noted. Statistical analysis was performed using Student's *t*-test for comparisons between two groups, one-way ANOVA with multiple comparisons for assessment of more than two groups on GraphPad Prism software. Comparisons among specific groups were done using post-tests as indicated in the respective figure legends.



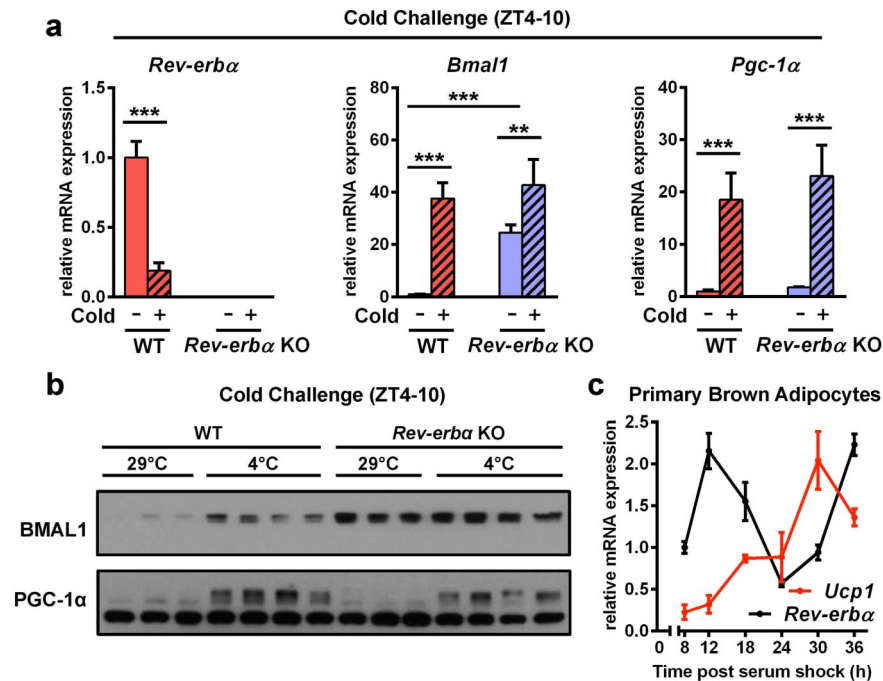
Extended Data Figure 2 | *Rev-erb α* controls cold and noradrenaline-induced oxidative metabolism independently of skeletal muscle metabolism. **a**, Food intake from cold-challenged *Rev-erb α* knockout mice and control littermates in Fig. 1f. **b**, r.m.s. derivation of EMG measurement from Fig. 1g. **c**, Oxygen consumption rates of *Rev-erb α* KO mice and control littermates following noradrenaline administration ($1 \text{ mg kg}^{-1} \text{ s.c.}$) ($n = 6$). **d**, **e**, r.m.s. derivation of EMG measurements performed on wild-type and *Rev-erb α* knockout mice following noradrenaline administration ($1 \text{ mg kg}^{-1} \text{ s.c.}$) ($n = 4$). *** $P < 0.001$ as determined by Student's *t*-test. Data are expressed as mean \pm s.d.

Extended Data Figure 1 | The BAT core clock is largely unaffected by *Rev-erb α* deletion. **a**, *Rev-erb α* protein levels in BAT of wild-type and *Rev-erb α* knockout mice ($n = 2$; each lane of the western blot represents pooled biological duplicates). **b**, BAT mRNA for indicated genes from wild-type and *Rev-erb α* knockout mice collected at the indicated times over a 24-h time course ($n = 3$).



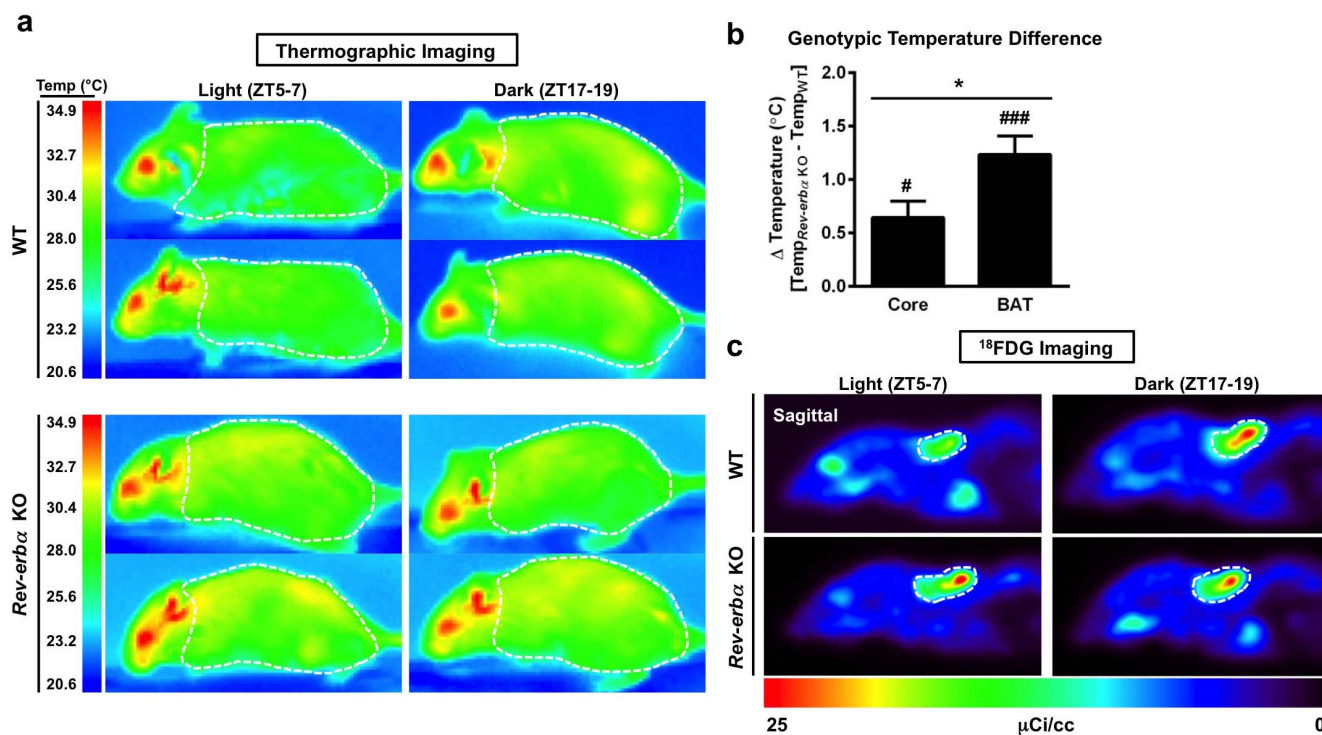
Extended Data Figure 3 | *Rev-erbα*, but not *Rev-erbβ*, is decreased in a cold-dependent manner. **a–c**, *Rev-erbβ* (**a**), *Bmal1* (**b**) and *Pgc1α* (**c**) mRNA levels in BAT during a cold-exposure time course ($n = 3$ for mRNA). **d**, BAT gene expression following moderate (20°C) or acute (4°C) cold challenges

($n = 3$). **e**, BAT protein levels after 3 h noradrenaline administration ($1 \text{ mg kg}^{-1} \text{ i.p.}$) or cold exposure ($n = 3$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ as determined by one-way ANOVA with multiple comparisons and a Tukey's post-test. Data are expressed as mean \pm s.d.



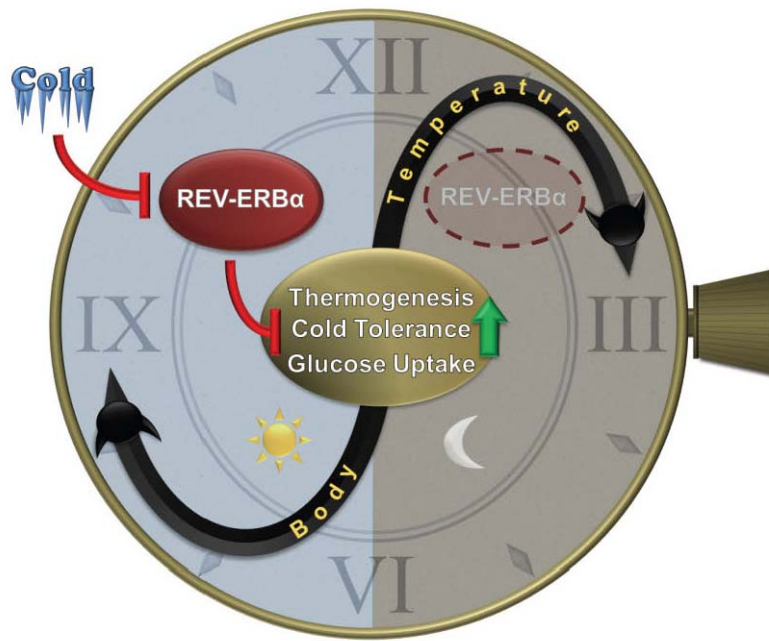
Extended Data Figure 4 | *Rev-erbα* negatively regulates *Ucp1*. a, b, BAT mRNA (a) and protein (b) from wild-type and *Rev-erbα* knockout mice exposed to cold for 6 h as described in Fig. 3a, b. c, mRNA levels in preadipocytes isolated from wild-type mice, differentiated in culture and

collected at the indicated times after synchronization by serum shock ($n = 4$). ** $P < 0.01$, *** $P < 0.001$ as determined by one-way ANOVA with multiple comparisons and a Tukey's post-test. Data are expressed as mean \pm s.d.



Extended Data Figure 5 | *Rev-erb α* controls circadian oscillation of surface temperature and BAT activity. **a**, Infrared images from the thermographic surface temperature analysis performed in Fig. 4c. **b**, Genotypic differences between BAT and core temperatures from wild-type and *Rev-erb α* knockout mice acclimated to thermoneutrality ($n = 6$). **c**, ¹⁸FDG imaging ($n = 4$) of *Rev-erb α* knockout mice and wild-type littermates during the light and dark

phases. Representative sagittal planes are shown for each group. * $P < 0.05$, Δ Core temperature versus Δ BAT temperature; † $P < 0.05$, core temperature versus *Rev-erb α* knockout core temperature; ‡ $P < 0.001$, wild-type BAT temperature versus *Rev-erb α* knockout BAT temperature as determined by Student's *t*-test. Data are expressed as mean \pm s.e.m.



Extended Data Figure 6 | The nuclear receptor Rev-erb α controls circadian thermogenic plasticity. Rev-erb α regulates the circadian rhythm of body temperature through direct suppression of thermogenesis and BAT activity.

Cold exposure during the light phase rapidly overrides Rev-erb α -dependent repression to induce thermogenic programs.

Temperature-dependent regulation of flowering by antagonistic FLM variants

David Pose^{1†}, Leonie Verhage^{2,3}, Felix Ott¹, Levi Yant^{1†}, Johannes Mathieu^{1†}, Gerco C. Angenent^{2,3}, Richard G. H. Immink² & Markus Schmid¹

The appropriate timing of flowering is crucial for plant reproductive success. It is therefore not surprising that intricate genetic networks have evolved to perceive and integrate both endogenous and environmental signals, such as carbohydrate and hormonal status, photoperiod and temperature^{1,2}. In contrast to our detailed understanding of the vernalization pathway, little is known about how flowering time is controlled in response to changes in the ambient growth temperature. In *Arabidopsis thaliana*, the MADS-box transcription factor genes *FLOWERING LOCUS M* (*FLM*) and *SHORT VEGETATIVE PHASE* (*SVP*) have key roles in this process^{3,4}. *FLM* is subject to temperature-dependent alternative splicing³. Here we report that the two main *FLM* protein splice variants, *FLM-β* and *FLM-δ*, compete for interaction with the floral repressor *SVP*. The *SVP-FLM-β* complex is predominately formed at low temperatures and prevents precocious flowering. By contrast, the competing *SVP-FLM-δ* complex is impaired in DNA binding and acts as a dominant-negative activator of flowering at higher temperatures. Our results show a new mechanism that controls the timing of the floral transition in response to changes in ambient temperature. A better understanding of how temperature controls the molecular mechanisms of flowering will be important to cope with current changes in global climate^{5,6}.

Distinct aspects of temperature contribute to flowering time control. The vernalization pathway controls flowering of winter-annual *Arabidopsis* accessions in response to prolonged periods of cold by the epigenetic silencing of the potent floral repressor *FLOWERING LOCUS C* (*FLC*)^{7–9}. Ambient temperature also has an essential role, inducing flowering in *Arabidopsis* at warmer temperatures under otherwise non-inductive short-day photoperiods³. The identification of mutants affected in the ambient temperature response indicates that there is a strong genetic contribution to the regulation of flowering in response to temperature changes¹. Most recently, *H2A.Z* and *PHYTOCHROME INTERACTING FACTOR4* (*PIF4*) have emerged as important positive regulators of flowering in response to temperature, the latter being essential for the induction of flowering in response to warmer temperatures in short-day conditions^{10–12}. By contrast, the MIKC-type MADS-domain transcription factors *SVP*⁴, *FLM*³ and *MADS AFFECTING FLOWERING* 2, 3 and 4 (*MAF2–4*; also known as *AGL31*, *AGL70* and *AGL69*, respectively)^{13,14} have been described as negative regulators of flowering. Both *SVP* and *FLM* contribute to flowering-time variation between natural accessions of *Arabidopsis*^{15,16}, and mutations in these genes lead to decreased thermosensitivity^{3,4,14} (Table 1, experiment 1). Moreover, *SVP* and *FLM* have been shown to interact genetically¹⁷. All together, these findings suggest a role for these two MADS transcription factors in the ambient temperature pathway. The *FLM* transcript is subject to alternative splicing, with four splice variants (α , β , γ and δ) expressed in the Wassilewskija accession¹⁸. The *Arabidopsis* Columbia (Col-0) accession primarily transcribes two splice variants, *FLM-β* and *FLM-δ*

(Extended Data Fig. 1a, b), which are both translated¹⁹. They incorporate either the second or third exon, respectively, which encodes part of the MIKC ‘intervening’ (I) region that is thought to contribute to protein–protein interaction properties²⁰. Interestingly, *FLM* splicing changes in response to ambient temperature variation³, suggesting that the proteins encoded might affect flowering in different ways.

To understand the effect of temperature-dependent alternative splicing of *FLM* on flowering we analysed the effect of ambient temperature fluctuation on *FLM* splicing. *FLM-β* and *FLM-δ* were detected at similar ratios in all tissues examined (Fig. 1a, b). By contrast, expression of the two transcripts was different in plants that had been grown

Table 1 | Flowering times of mutants and transgenic plants

		RLN	CLN	TLN	s.d.	Range	<i>n</i>
Experiment 1 (LD)							
Col-0	16 °C	17.3	5.1	22.5	±1.8	19–26	20
	23 °C	12.6	2.6	15.2	±1.6	13–17	19
	27 °C	9.5	2.7	12.2	±1.1	10–14	29
<i>flm-3</i>	16 °C	9.9	3.4	13.3	±1.2	11–16	21
	23 °C	10.0	3.0	13.0	±1.0	11–14	18
	27 °C	8.5	2.8	11.3	±1.3	8–14	34
<i>svp-32</i>	16 °C	6.5	2.6	9.1	±0.7	8–11	22
	23 °C	7.2	3.4	10.5	±0.9	9–12	25
	27 °C	6.7	2.8	9.4	±0.9	8–11	19
<i>svp-32/flm-3</i>	16 °C	6.9	2.7	9.6	±1.2	8–11	20
	23 °C	6.8	3.1	10.0	±1.0	8–12	23
	27 °C	6.9	2.8	9.9	±1.4	8–12	28
Experiment 2 (16 °C LD)							
Col-0 BAR		18.1	3.9	21.9	±2.9	15–27	29
<i>flm-3</i> BAR		10.7	3.5	14.2	±1.7	11–17	27
35S: <i>FLM-β</i> #11		23.8	9.7	33.4	±5.9	23–40	12
35S: <i>FLM-β</i> #21		26.9	11.0	37.9	±3.1	33–42	9
35S: <i>FLM-δ</i> #1		12.0	3.6	15.6	±2.2	13–20	13
35S: <i>FLM-δ</i> #4		12.9	3.4	16.4	±1.5	14–19	14
<i>flm-3</i> 35S: <i>FLM-β</i> #39		17.5	5.2	22.8	±2.5	19–28	21
<i>flm-3</i> 35S: <i>FLM-β</i> #54		17.4	5.2	22.6	±3.6	17–29	15
<i>flm-3</i> 35S: <i>FLM-δ</i> #3		8.9	2.8	11.7	±1.4	8–14	28
<i>flm-3</i> 35S: <i>FLM-δ</i> #43		8.8	2.4	11.3	±2.1	7–14	28
Experiment 3 (16 °C LD)							
Col-0 BAR		19.9	5.8	25.7	±2.9	20–30	24
<i>flm-3</i> BAR		8.9	3.2	12.1	±1.6	9–14	14
<i>flm-3</i> pFLM:gFLM #2		21.2	7.0	28.2	±3.5	21–32	13
<i>flm-3</i> pFLM:gFLM #3		21.0	6.8	27.8	±2.4	25–31	5
<i>flm-3</i> pFLM:gFLM-GFP #2		19.4	7.6	27.0	±2.2	22–30	14
<i>flm-3</i> pFLM:gFLM-GFP #4		19.9	7.4	27.3	±2.2	25–32	11
Experiment 4 (16 °C LD)							
Col-0 BAR		22.0	4.8	26.8	±2.7	21–31	26
<i>flm-3</i> BAR		13.7	3.7	17.4	±1.5	14–20	29
<i>flm-3</i> pFLM:iFLM- <i>β</i> #24		26.0	6.8	32.8	±3.2	28–39	18
<i>flm-3</i> pFLM:iFLM- <i>δ</i> #17		11.9	3.4	15.4	±1.6	12–19	29
<i>flm-3</i> pFLM:iFLM- <i>β</i> -GFP #10		20.6	5.6	26.2	±1.9	20–32	40
<i>flm-3</i> pFLM:iFLM- <i>δ</i> -GFP #8		12.5	3.6	16.1	±1.3	13–19	40

n denotes the number of individuals; # denotes the identifier of individual transgenic lines. BAR, glufosinate (BASTA) resistance; CLN, cauline leaf number; LD, long-day; RLN, rosette leaf number; s.d., total leaf number standard deviation; TLN, total leaf number.

¹Max Planck Institute for Developmental Biology, Department of Molecular Biology, Spemannstr. 35, 72076 Tübingen, Germany. ²Plant Research International, Bioscience, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands. ³Laboratory of Molecular Biology, Wageningen University, 6708 PB Wageningen, The Netherlands. [†]Present addresses: Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga–Consejo Superior de Investigaciones Científicas, Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain (D.P.); Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA (L.Y.); Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, New York 14853-1801, USA (J.M.).

at different temperatures (Fig. 1c). *FLM-β* was the prevalent splice variant at 16 °C, whereas *FLM-δ* dominated at 27 °C (Fig. 1c). In addition, after a shift from 16 °C to 27 °C, expression of *FLM-β* decreased within 24 h, whereas *FLM-δ* levels increased (Fig. 1d). The opposite was the case when plants were transferred from 27 °C to 16 °C (Fig. 1d). By contrast, *SVP* expression was only moderately induced with increasing temperature (Fig. 1e). Because *SVP* regulates flowering in response to ambient temperature, this suggests that *SVP* might be regulated at the post-transcriptional or protein–protein interaction level⁴.

To investigate the effect of the *FLM* splice variants on flowering time, we expressed the *FLM-β* and *FLM-δ* open-reading frames (ORFs) from the 35S promoter in wild-type and *flm-3* plants. As expected for a floral repressor, expression of *FLM-β* strongly delayed flowering in Col-0, and complemented the early flowering of *flm-3* (Table 1, experiment 2, and Extended Data Fig. 2a). *FLM-δ* expression had the opposite effect and induced early flowering (Table 1, experiment 2, and Extended Data Fig. 2a). Expression analysis confirmed that endogenous *FLM-β* and *MAF2*, *MAF3*, *MAF5* (also known as *AGL68*) and *FLC*, as well as *SVP*, were expressed normally in 35S:*FLM-δ* Col-0 plants, indicating that the early flowering phenotype was not caused by co-suppression (Extended Data Fig. 3a, b). In addition, crosses between 35S:*FLM-β* and 35S:*FLM-δ* plants displayed an intermediate phenotype (Extended Data Fig. 3c–e), suggesting that *FLM-δ* is responsible for the acceleration of flowering of these lines.

FLM-β, but not *FLM-δ*, was able to form homodimers, and the two proteins were able to heterodimerize, according to yeast two-hybrid analyses (Fig. 2a and Supplementary Table 1). Furthermore, both *FLM* isoforms interacted with fully spliced *SVP* (At2g22540.1) but not with any other variant tested (Supplementary Table 1). All interactions were

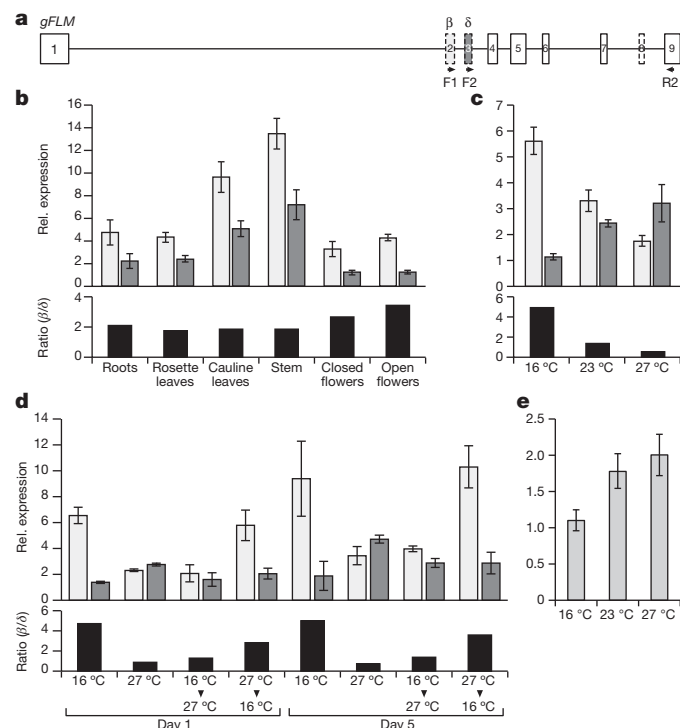


Figure 1 | Temperature-dependent expression of *FLM-β*, *FLM-δ* and *SVP*. **a**, *FLM* locus, including exons (boxes) and introns (lines). Primers used for qRT-PCR are indicated. **b–d**, Relative (rel.) expression of *FLM-β* (light grey) and *FLM-δ* (dark grey) in different tissues in 10-day-old Col-0 seedlings grown at 23 °C long-day (**b**), in whole seedlings at 16 °C, 23 °C and 27 °C long-day (**c**) and days 1 and 5 after a shift between temperatures (**d**). The ratio of *FLM-β*/*FLM-δ* expression is shown in black. **e**, *SVP* expression in 10-day-old Col-0 seedlings grown at 16 °C, 23 °C and 27 °C. Error bars denote s.d. of three biological replicates with three technical repetitions each.

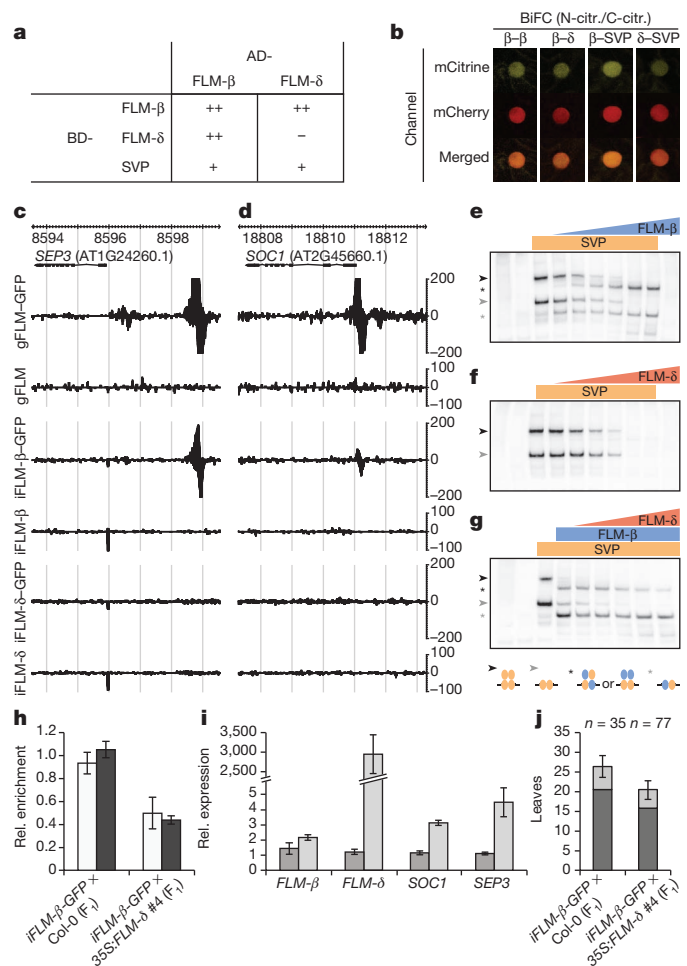


Figure 2 | *FLM*-*SVP* protein–protein interactions and DNA binding assays. **a**, Yeast two-hybrid. AD, activation domain; BD, DNA-binding domain. **b**, BiFC *FLM-β*, *FLM-δ* and *SVP* interaction assays. C-citr., C-terminal half of haemagglutinin (HA)-tagged mCitrine; N-citr., N-terminal half of Myc-tagged mCitrine. **c**, **d**, Local enrichment of GFP-tagged *FLM*, *iFLM-β* and *iFLM-δ* bound to the *SEP3* (**c**) and *SOC1* (**d**) regulatory regions assayed by ChIP-seq. Each panel shows a 6-kilobase (kb) window. **e–g**, EMSA competition assays using a *SEP3* promoter probe containing two CarG motifs. Lanes 1 and 2 correspond to ‘no protein’ and ‘shuffled-*SVP*’ controls, respectively. Increasing concentrations of *FLM-β* (**e**) and *FLM-δ* (**f**) were added to a constant amount of *SVP*. **g**, Titration of *FLM-δ* to constant amounts of *SVP* and *FLM-β*. Orange and blue ellipses represent *SVP* and *FLM-β* proteins, respectively. **h**, Relative enrichment of binding of *iFLM-β*-GFP to the promoters of *SOC1* (open bars) and *SEP3* (filled bars) in *iFLM-β*-GFP × 35S:*FLM-δ* F₁ and control plants. **i**, Expression of *SOC1* and *SEP3* in *iFLM-β*-GFP × 35S:*FLM-δ* F₁ (light grey) and control (dark grey) plants. Error bars denote the s.d. of three biological replicates with three technical repetitions each. **j**, Flowering time of the F₁ plants. Rosette and cauline leaf number are represented in dark and light grey.

confirmed by transient bimolecular complementation (BiFC) assays (Fig. 2b and Extended Data Fig. 4). In addition, both *FLM* isoforms also interacted in yeast with the type-I MADS-domain protein *AGL74N* (At1g48150; Supplementary Table 1). However, an *AGL74N* T-DNA insertion allele flowered normally (Extended Data Fig. 5). The finding that both *FLM-β* and *FLM-δ* were able to interact with the floral repressor *SVP*, but had opposite effects on flowering time, suggests a model in which the incorporation of a particular *FLM* isoform determines the activity of the resulting *SVP*-*FLM* heterocomplex.

To test this model and identify *FLM* direct targets we constructed a genomic *FLM* (*gFLM*) fragment that rescued *flm-3*, independently of whether a carboxy-terminal enhanced variant green fluorescent protein tag (mGFP6) is present (Table 1, experiment 3, and Extended Data Fig. 2b).

Chromatin immunoprecipitation and massively parallel sequencing (ChIP-seq) on a rescued gFLM-GFP line revealed binding of FLM to the regulatory regions of several flowering time related genes, including *SUPPRESSOR OF CONSTANS OVEREXPRESSION 1* (*SOC1*; also known as *AGL20*), *Arabidopsis thaliana* *CENTRORADIALIS* homologue (*ATC*), *TEMPRANILLO 2* (*TEM2*; also known as *RAV2*) and *SCHLAFLMUTZE* (*SMZ*), and floral homeotic genes such as *SEPALATA 3* (*SEP3*), *APETALA 3* (*AP3*) and *PISTILLATA* (*PI*) (Fig. 2c, d, Extended Data Fig. 6b–e and Supplementary Table 2).

To identify splice-variant-specific targets of FLM, we established transgenic lines that expressed *FLM-β* (*pFLM:iFLM-β*) or *FLM-δ* (*pFLM:iFLM-δ*), with or without the C-terminal mGFP tag, in *flm-3* (Table 1, experiment 4, and Extended Data Figs 2c and 6a–e). ChIP-seq performed using an *iFLM-β-GFP* rescue line revealed that most of the targets (67%) identified in the *gFLM-GFP* line were bound by *iFLM-β-GFP* (Supplementary Tables 3 and 4 and Extended Data Fig. 7a). Quantitative reverse transcriptase PCR (qRT-PCR) analysis of *SEP3*, *SOC1* and *TEM2* confirmed that these genes were also regulated in their expression by FLM-β (Extended Data Fig. 6f–h). In addition, expression of an FLM-β-VP16 fusion protein (containing the VP16 transcriptional activation domain) resulted in early flowering (Extended Data Fig. 2d), indicating that FLM-β delays flowering mainly through transcriptional repression. *SEP3*, *SOC1* and *TEM2* were also found among 61 genes that were shared between our *iFLM-β* ChIP-seq data and SVP targets identified by ChIP coupled to DNA microarray (ChIP-chip)²¹ (Supplementary Table 4 and Extended Data Fig. 7b), suggesting that these genes could be regulated by an SVP–FLM-β heterocomplex. By contrast, *iFLM-δ-GFP* ChIP resulted in only minor enrichment at few loci (Supplementary Table 3), suggesting that FLM-δ does not bind to DNA efficiently.

To analyse the *in vitro* DNA-binding properties of SVP, FLM-β and FLM-δ in detail, we performed electrophoretic mobility shift assays (EMSAs). We observed strong binding of SVP to CA₂G boxes in the same regions of the *SEP3*, *SOC1* and *ATC* promoters that had shown enrichment for FLM and SVP binding in ChIP-seq and ChIP-chip analyses, respectively (Extended Data Fig. 8a–c). By contrast, no changes in DNA mobility were observed for FLM-δ and FLM-β (Extended Data Fig. 8a–c), indicating that these two proteins alone do not bind DNA *in vitro*. However, simultaneous *in vitro* transcription/translation

of SVP and FLM-β, but not of SVP and FLM-δ, resulted in additional bands, corresponding to different SVP–FLM-β heterocomplexes (Fig. 2e and Extended Data Fig. 8a–c). These findings indicate that FLM-β requires SVP for DNA binding *in vitro*. In agreement with this, we found that the late flowering of *35S:FLM-β* plants was abolished in the *svp-32* mutant (Fig. 3b), suggesting that FLM-β depends on SVP to repress flowering. This dependency seems to be bidirectional, as the late flowering and the homeotic transformation of sepals and petals characteristic of an SVP mis-expression line are suppressed in *flm-3* (Fig. 3a, b). Moreover, a double *flm-3* *svp-32* mutant line did not display any additional phenotype to that of the single *svp-32* single mutant¹⁷ (Fig. 3b), and a *35S:FLM-β* *35S:SVP* double-transgenic line flowered much later than the individual mis-expression lines (Fig. 3d).

To identify the molecular mechanism underlying the flowering-promoting effect of FLM-δ, we performed EMSA competition experiments in which increasing amounts of FLM-δ were added to either SVP (Fig. 2f) or SVP plus FLM-β (Fig. 2g). We observed an FLM-δ concentration-dependent reduction in DNA binding for both SVP and SVP–FLM-β (Fig. 2f, g and Extended Data Fig. 8d, e), suggesting that FLM-δ functions as a dominant-negative version of FLM that alleviates SVP–SVP and SVP–FLM-β mediated repression by replacing one of the interaction partners in the complex, thereby rendering it inactive. As predicted by this model, constitutive expression of *FLM-δ* suppressed the late flowering and flower phenotypes of *35S:SVP* plants (Fig. 3c, d). Furthermore, binding of *iFLM-β-GFP* to the promoters of *SEP3* and *SOC1* was reduced in *pFLM:iFLM-β-GFP* × *35S:FLM-δ* plants, indicating that FLM-δ negatively affects the ability of FLM-β to bind to its target (Fig. 2h). In addition, expression of these two genes was increased (Fig. 2i) in the double transgenic line, which also flowered considerably earlier than control plants (Fig. 2j).

Our results demonstrate that the protein isoforms encoded by two splice variants of *FLM*, *FLM-β* and *FLM-δ*, compete for interaction with SVP to regulate flowering in opposition. Low ambient temperature favours the formation of SVP–SVP and SVP–FLM-β complexes that actively repress flowering. However, as temperatures increase, not only is the amount of the FLM-β spliceform downregulated, but the flower-repressive function of SVP and the remaining FLM-β proteins are counteracted by a relative increase in the dominant-negative regulator of flowering, FLM-δ (Extended Data Fig. 9). We propose that the

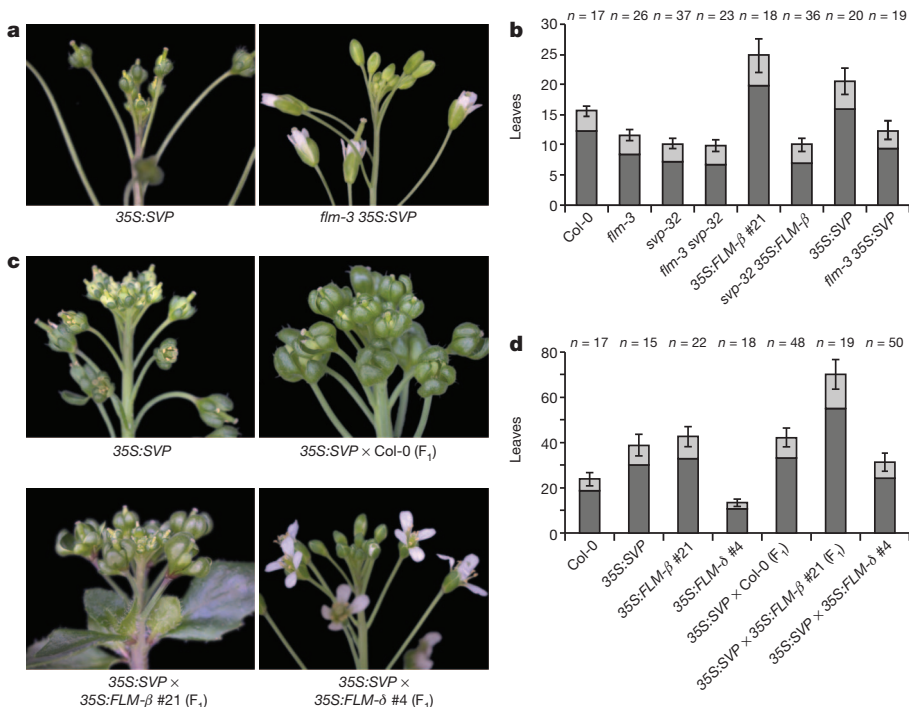


Figure 3 | FLM and SVP are interdependent and regulate flowering time and flower morphology. **a**, *flm-3* suppresses the green and leaf-like sepal and petal phenotype of *35S:SVP* flowers. **b**, Flowering time of *flm-3* and *svp-32* single and double mutants, and FLM-β and SVP mis-expression lines in wild-type (Col-0), *svp-32* or *flm-3*. **c**, *35S:FLM-δ* suppresses the *35S:SVP* flower phenotype. **d**, *35S:FLM-β* enhances the late flowering of *35S:SVP*, whereas *35S:FLM-δ* has the opposite effect. Plants were grown at 23 °C (**a**, **b**) and 16 °C (**c**, **d**). Rosette and cauline leaf number are represented in dark and light grey, respectively.

opposing activities of two different splice variants of *FLM*, and possibly other transcription factors that are also subject to (temperature-dependent) alternative splicing^{13,22–24}, constitute a distinct regulatory pathway that acts in parallel to PIF4 to reinforce the transition from vegetative growth to reproductive development in response to changes in ambient temperature. It will be interesting to see how additional regulatory mechanisms such as posttranslational modification or targeted degradation of *FLM* or *SVP* contribute to the regulation of flowering in response to changes in ambient temperature. Furthermore, given that natural variation in *FLM* contributes substantially to flowering time differences among *Arabidopsis* accessions^{3,15}, the role of temperature-dependent mRNA splicing in adaptation to climate change is worthy of special focus.

METHODS SUMMARY

Plant material and growth conditions. Plants were grown in chambers in long-day conditions (16 h light/8 h dark) at 16 °C, 23 °C and 27 °C.

ChIP-seq experiments. Two biological replicates were performed for all the ChIP assays except for the *flm-3 pFLM:gFLM* and *flm-3 pFLM:gFLM-GFP* lines, for which three biological replicates were assayed. DNA was fragmented and precipitated. Libraries for high-throughput sequencing were prepared and 40-base-pair single-end sequencing was performed on an Illumina GAIIx instrument. ChIP-seq peak calling was performed using the SHORE software version 0.8 (<http://1001genomes.org/software/shore.html>).

EMSA assays. Oligonucleotide sequences of the 5′-biotinylated probes used were determined based on the *FLM*-binding sites identified by ChIP-seq. To establish protein–protein interactions, transcription and translation were performed in a single tube reaction for each protein complex.

Yeast two-hybrid assays. A matrix-based interaction assay was performed, in which *FLM-β* and *FLM-δ* were screened against the full collection of *Arabidopsis* MADS domain transcription factors.

BiFC analyses. The vectors were co-infiltrated into 3-week-old tobacco (*Nicotiana benthamiana*) leaves. Competition assays were performed to reduce false positives. The nuclear marker 35S::NLS-*mCherry* and the silencing suppressor 35S::*p19* were included in all infiltrations.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 July; accepted 5 September 2013.

Published online 25 September 2013.

1. Srikanth, A. & Schmid, M. Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* **68**, 2013–2037 (2011).
2. Posé, D., Yant, L. & Schmid, M. The end of innocence: flowering networks explode in complexity. *Curr. Opin. Plant Biol.* **15**, 45–50 (2012).
3. Balasubramanian, S., Sureshkumar, S., Lempe, J. & Weigel, D. Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genet.* **2**, e106 (2006).
4. Lee, J. H. *et al.* Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. *Genes Dev.* **21**, 397–402 (2007).
5. Marcott, S. A., Shakun, J. D., Clark, P. U. & Mix, A. C. A reconstruction of regional and global temperature for the past 11,300 years. *Science* **339**, 1198–1201 (2013).
6. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105**, 13252–13257 (2008).

7. Chouard, P. Vernalization and its relations to dormancy. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **11**, 191–238 (1960).
8. Sheldon, C. C. *et al.* The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**, 445–458 (1999).
9. Michaels, S. D. & Amasino, R. M. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
10. Kumar, S. V. & Wigge, P. A. H2A-Z-containing nucleosomes mediate the thermosensory response in *Arabidopsis*. *Cell* **140**, 136–147 (2010).
11. Kumar, S. V. *et al.* Transcription factor PIF4 controls the thermosensory activation of flowering. *Nature* **484**, 242–245 (2012).
12. Franklin, K. A. *et al.* Phytochrome-interacting factor 4 (PIF4) regulates auxin biosynthesis at high temperature. *Proc. Natl Acad. Sci. USA* **108**, 20231–20235 (2011).
13. Rosloski, S. M. *et al.* Functional analysis of splice variant expression of *MADS AFFECTING FLOWERING 2* of *Arabidopsis thaliana*. *Plant Mol. Biol.* **81**, 57–69 (2013).
14. Gu, X. *et al.* *Arabidopsis* FLC clade members form flowering-repressor complexes coordinating responses to endogenous and environmental cues. *Nature Commun.* **4**, 1947 (2013).
15. Werner, J. D. *et al.* Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. *Proc. Natl Acad. Sci. USA* **102**, 2460–2465 (2005).
16. Méndez-Vigo, B., Martínez-Zapater, J. M. & Alonso-Blanco, C. The flowering repressor *SVP* underlies a novel *Arabidopsis thaliana* QTL interacting with the genetic background. *PLoS Genet.* **9**, e1003289 (2013).
17. Scortecci, K., Michaels, S. D. & Amasino, R. M. Genetic interactions between *FLM* and other flowering-time genes in *Arabidopsis thaliana*. *Plant Mol. Biol.* **52**, 915–922 (2003).
18. Scortecci, K. C., Michaels, S. D. & Amasino, R. M. Identification of a MADS-box gene, *FLOWERING LOCUS M*, that represses flowering. *Plant J.* **26**, 229–236 (2001).
19. Jiao, Y. & Meyerowitz, E. M. Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* **6**, 419 (2010).
20. Riechmann, J. L., Krizek, B. A. & Meyerowitz, E. M. Dimerization specificity of *Arabidopsis* MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proc. Natl Acad. Sci. USA* **93**, 4793–4798 (1996).
21. Tao, Z. *et al.* Genome-wide identification of *SOC1* and *SVP* targets during the floral transition in *Arabidopsis*. *Plant J.* **70**, 549–561 (2012).
22. James, A. B. *et al.* Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* **24**, 961–981 (2012).
23. Wang, X. *et al.* *SKIP* is a component of the spliceosome linking alternative splicing and the circadian clock in *Arabidopsis*. *Plant Cell* **24**, 3278–3295 (2012).
24. Jones, M. A. *et al.* Mutation of *Arabidopsis* spliceosomal timekeeper locus1 causes circadian clock defects. *Plant Cell* **24**, 4066–4082 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Hohnstein, J. Weirich, C. Martín-Pizarro, J. Busscher-Lange, C. Lanz and J. Reixinger for technical assistance, A. Srikanth for BiFC vectors, and P. Huijser and B. Davies for seeds. Supported through the ERA-NET PLANT GENOMICS collaborative project BLOOM-NET by a grant from the Deutsche Forschungsgemeinschaft (SCHM 1560/7-1) and the MPI for Developmental Biology to M.S., a BLOOM-NET related NWO grant to R.G.H.I. and G.C.A., and a NWO/TTI-Green Genetics THERMOFLOW grant to R.G.H.I., G.C.A. and L.V.

Author Contributions D.P. performed most experiments, L.V. performed EMSA assays, F.O. analysed ChIP-seq data, L.Y. and J.M. generated 35S::*FLM-β*/35S::*FLM-δ* constructs, L.Y. generated and phenotyped F₁ Col-0 35S::*FLM-β*/35S::*FLM-δ* transgenic plants, G.C.A. contributed to the discussion, and D.P., R.G.H.I., L.Y. and M.S. supervised the work and wrote the manuscript.

Author Information ChIP-seq data have been deposited with NCBI Gene Expression Omnibus under accession number GSE48082. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (markus.schmid@tuebingen.mpg.de).

METHODS

Plant material and growth conditions. All plants used in this work are of the Columbia (Col-0) accession. The *flm-3* and *svp-32* lines have been described previously (Extended Data Table 1). Genotypes were confirmed by PCR using oligonucleotides listed in Supplementary Table 5. Plants were grown in chambers in a long-day condition (16 h light/8 h dark) controlled photoperiod at 16 °C, 23 °C and 27 °C, 65% humidity and a mixture of Cool White and Gro-Lux Wide Spectrum fluorescent lights, with a fluence rate of 125–175 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

Plasmid constructions and plant transformation. All oligonucleotides used in this work are listed in Supplementary Table 5. All PCR reactions were performed using Phusion polymerase (New England Biolabs) and the constructs were verified by Sanger sequencing after cloning.

The ORFs of *FLM* (At1g77080; different splice forms) and *SVP* (At2g2540.1) were amplified from cDNA prepared from Col-0 seedlings grown at 23 °C using primers G-2196 and G-1978, and primers G-28863 and G-28864, respectively. The PCR product was purified and ligated into the pJLSmart Gateway entry vector, generating the constructs *pJM313* (*FLM-β*), *pLY224* (*FLM-δ*) and *pDP57* (*SVP*). pJLSmart is a modified pENTR1A vector (Invitrogen, Life Technologies) with the polylinker and *ccdB* gene replaced by a minimal MCS, flanked by attL1 and attL2 sites, to allow for blunt-end cloning using a SmaI site and subsequent Gateway recombination. For mis-expression experiments, the ORFs were introduced into the Gateway-compatible pGREEN-IIS binary destination vectors *pFK210* (*FLM-β* and *FLM-δ*) and *pFK209* (*SVP*), which provide resistance to BASTA and kanamycin, respectively, for selection in plants, by Gateway recombination, resulting in *pJM356* (35S:*FLM-β*), *pLY225* (35S:*FLM-δ*) and *pDP58* (35S:*SVP*). For the yeast two-hybrid assays the ORFs were recombined into the destination vectors *pDEST22* (AD) and *pDEST32* (BD), creating *pDP10* (AD-*FLM-β*), *pDP11* (BD-*FLM-β*), *pDP14* (AD-*FLM-δ*) and *pDP15* (BD-*FLM-δ*).

For the BiFC assays, the ORFs were amplified without the stop codon using the oligonucleotides G-2196 and G-31569 for the *FLM-β* and *FLM-δ* splice variants, and G-28863 and G-31886 for *SVP*. The products were cloned into the pCR8/GW/TOPO Gateway entry vector (Invitrogen, Life Technologies) to create *pDP105* (*FLM-β* Δstop), *pDP106* (*FLM-δ* Δstop) and *pDP129* (*SVP* Δstop). The ORFs were recombined into destination vectors *pAS057* (Myc-mCitrine N-terminal) and *pAS061* (HA-mCitrine C-terminal) to generate the split-mCitrine-tagged constructs for *FLM-β* (*pDP157* and *pDP158*), *FLM-δ* (*pDP159* and *pDP160*) and *SVP* (*pDP161* and *pDP162*).

For the 35S:*FLM-β*-VP16 construct, we recombined the *FLM-β* ORF from *pDP105* into the destination vector *pFK250*, which provides an in-frame VP16 tag, resulting in the construct *pDP175*.

The 6,876-bp genomic *FLM* (At1g77080; TAIR10, chr1: 28953510..28960386) rescue fragment, which includes approximately 2.1 kb upstream sequence, exons, introns, the 3' untranslated region and approximately 0.3 kb downstream sequence, was amplified by PCR from genomic DNA isolated from Col-0 using the primers G-26819 and G-26820 (Supplementary Table 5). The resulting PCR product was purified and cloned into the pCR8/GW/TOPO vector to create *pDP22*. Subsequently, the *FLM* genomic fragment was recombined from *pDP22* into the pGREEN-IIS binary destination vector *pFK387*, which provides resistance to BASTA for selection in plants, resulting in *pDP34* (*pFLM:gFLM*). To facilitate ChIP, *FLM* was tagged with mGFP6-6×His. For this purpose, we amplified a genomic sub-fragment of *FLM* ranging from exon 7, which contains a unique SexAI restriction site, to the last coding triplet before the stop codon of *FLM* using the primers G-22798/G-26831, and the *FLM* 3' region, which contains a unique SacI restriction site, starting with the stop codon using the primers G-26335 and G-26820. The sequence encoding mGFP6-6×His was amplified from plasmid pMD107 (ref. 25) using the primers G-26832 and G-26334. Next, the three fragments were combined in an overlapping fusion PCR using primers G-22798 and G-26820. The resulting PCR product (5'-*FLM-mGFP6-3'*; Supplementary Table 5) was cloned into the pGEM-T Easy vector (Promega) to create *pDP23* and subsequently cut with SexAI and SacI and cloned into the corresponding sites of *pDP22* to generate *pDP24*. Finally, the *FLM-mGFP6* genomic fragment was recombined into *pFK387* to create *pDP28* (*pFLM:gFLM-GFP*).

The ORFs of the *FLM-β* and *FLM-δ* splice variants under control of the *FLM* 5' and 3' regions present in the *gFLM* construct were created by overlapping PCR. First, two halves were amplified separately with the oligonucleotides G-29612 and G-28155 using *pJM313* (*FLM-β*) and *pLY224* (*FLM-δ*) as templates, and with G-26118 and G-26820 using *pDP22* as template. The two halves were fused in a PCR using the oligonucleotides G-29612 and G-26820, purified and cloned into pGEM-T Easy to create *pDP67* (*FLM-β-3'*) and *pDP68* (*FLM-δ-3'*). These constructs were subsequently cut with NcoI and SacI and cloned into the corresponding sites of *pDP22* to generate *pDP75* (*pFLMp:FLM-β-3'*) and *pDP76* (*pFLMp:FLM-δ-3'*), which were recombined into *pFK387* to create *pDP79* and *pDP80*, respectively.

The first intron of *FLM* was introduced into the ORFs of *FLM-β* and *FLM-δ* by overlapping PCR. The 5' halves of the insert were amplified using the oligonucleotides G-29612 and G-30980 for *FLM-β*, and G-29612 and G-30982 for *FLM-δ*, using *pDP22* as template. The 3' halves were amplified using the oligonucleotides G-28150 and G-28154 for *FLM-β* using *pDP75* as a template, and with G-30981 and G-28154 for *FLM-δ* using *pDP76* as a template. The two fragments were fused in a PCR using the oligonucleotides G-29612 and G-28154, and the PCR product was purified and cloned into pGEM-T Easy to create *pDP92* (*FLM-β* fragment with first intron) and *pDP93* (*FLM-δ* fragment with first intron). These constructs were subsequently cut with NcoI and SexAI and cloned into the corresponding sites of *pDP75* and *pDP76* to generate *pDP94* (*pFLM:iFLM-β*) and *pDP95* (*pFLM:iFLM-δ*), respectively. These constructs were subsequently recombined into *pFK387* to create *pDP96* and *pDP97*, respectively. To perform ChIP, *pFLM:iFLM-β/δ* constructs were tagged with mGFP6-6×His by overlapping PCR. We first amplified with the oligonucleotides G-29345 and G-26831 a fragment of the *FLM* cDNA and with G-26832 and G-26820 the mGFP6-6×His plus the 3' region of *FLM* sequence. The two fragments were combined using the oligonucleotides G-29345 and G-26820 and cloned into pGEM-T Easy generating *pDP87* (*FLM* (7th–9th exon)-mGFP6-*FLM* 3'). *pDP87* was cut with SexAI and SacI and cloned into *pDP94* and *pDP95* creating *pDP101* (*pFLM:iFLM-β-GFP*) and *pDP102* (*pFLM:iFLM-δ-GFP*), respectively. The latter were recombined into the destination vector *pFK387* generating *pDP103* and *pDP104*, respectively.

All generated constructs were transformed into Col-0 and/or *flm-3* plants making use of *Agrobacterium tumefaciens* strain ASE and the floral dip method²⁶. Transgenic plants were identified by selective germination on soil watered with 0.1% glufosinate (BASTA).

RNA extraction, cDNA synthesis and expression analysis. For gene expression analysis, total RNA was isolated using TRIzol Reagent (Ambion, Life Technologies) according to manufacturer's instructions. One microgram of total RNA was DNase I-treated and single-stranded cDNA was synthesized using oligo(dT) and the RevertAid first-strand cDNA synthesis kit (Fermentas, Thermo Scientific). The resulting single-strand cDNA was diluted 25-fold and 4 μl was used as a template. Quantitative PCR (qPCR) was performed using the Platinum SYBR Green qPCR Supermix-UDG (Invitrogen, Life Technologies) and specific oligonucleotides (Supplementary Table 5) on an MJR Opticon Continuous Fluorescence Detection System. Relative expression values were calculated by the $\Delta\Delta C_t$ method using β -TUB2 (At5g62690) as a control. For each sample, material from a minimum of five seedlings was pooled. Error bars reported in Figs 1 and 2 and Extended Data Figs 3 and 6 denote the s.d. of three biological replicates with three technical repetitions each.

ChIP, library preparation and high throughput sequencing. ChIP was performed using 1 g of tissue collected at zeitgeber (ZT) 8 from homozygous *flm-3* *pFLM:gFLM* #2 and *flm-3* *pFLM:gFLM-GFP* #2 lines as well as from F₁ plants obtained from crosses of *flm-3* *pFLM:iFLM-β-GFP* #10 to Col-0 and 35S:*FLM-δ* #4, respectively, that were grown for 15 days under long days at 16 °C. ChIP on *flm-3* *pFLM:iFLM-β* #24, *flm-3* *pFLM:iFLM-β-GFP* #10, *flm-3* *pFLM:iFLM-δ* #17 and *flm-3* *pFLM:iFLM-δ-GFP* #8 was performed on seedlings grown for 10 long days at 23 °C.

Two biological replicates were performed for all the ChIP assays except for the *flm-3* *pFLM:gFLM* #2 and *flm-3* *pFLM:gFLM-GFP* #2 lines, for which three biological replicates were assayed. An anti-GFP antibody (Abcam; ab290) was used for immunoprecipitation. DNA was fragmented and precipitated as previously described²⁷. The resulting immunoprecipitated DNA was tested for enrichment by qPCR using presumed *FLM* targets such as *SEP3* and *SOC1* and a negative control locus from *ARR7* using primers described in the Supplementary Table 5. Libraries for high throughput sequencing were prepared as previously described²⁸ and 40-bp single-end sequencing was performed on an Illumina GAIIx instrument following the manufacturer's instructions.

ChIP-seq analyses. ChIP-seq peak calling was performed as previously described²⁹ using the SHORE software version 0.8 (shore.sf.net). The 40-bp reads were filtered and trimmed using the command 'shore import' with filtering and trimming options '-c -n 10% -k 32'. For mapping the reads to the TAIR10 reference sequence the 'shore mapflowcell' command was used using the GenomeMapper back end with a 10-mer reference sequence index. Alignment parameters were set as '-n 4-restrict = on', allowing for up to 4 base mismatches and no gaps. Peak calling was then performed on each pair of replicates using 'shore peak' with default peak calling parameters and option '-H 1,1' to exclude reads not assignable to a unique position on the reference sequence.

Matrix-based yeast two-hybrid assays. The pBD-GAL4 plasmids *pDP11* and *pDP15* were transformed into yeast strain PJ69-4A (mating type A³⁰) and the pAD-GAL4 *pDP10* and *pDP14* vectors into yeast strain PJ69-4α (mating type α). Three individual colonies from the pBD-GAL4 transformations were suspended in 50 μl sterile MQ water and spotted in 5 μl aliquots on synthetic dropout medium

lacking leucine and histidine and supplemented with 1, 3, 5 or 10 mM 3-amino-1,2,4-triazole. Growth of yeast and hence autoactivation, was scored after 5 days incubation at 20 °C, revealing that both FLM- β and FLM- δ do not have any autoactivation capacity. Subsequently, a matrix-based protein–protein interaction assay was performed as described previously³¹, in which FLM- β and FLM- δ were screened against the full collection of *Arabidopsis* MADS domain transcription factors³¹, extended by a collection of potential MADS domain protein isoforms³² and the SVP protein encoded by the fully spliced *SVP* gene³³. After mating the diploid yeast was spotted onto selective medium lacking leucine, tryptophan and histidine, and supplemented with 1 or 5 mM 3-amino-1,2,4-triazole. Growth of yeast was scored after 5 days incubation on the selective media at 20 °C. Three technical replicates were performed.

EMSA assays. ChIP-seq data for FLM (this article) was used to determine the oligonucleotide sequences of the probes. Probes were labelled with 5′-biotin either by cloning into pGEM-T vector and using vector-specific biotinylated primers (in case of the *SEP3* probe) or directly by sequence-specific biotinylated primers (in case of the *SOC1* and *ATC* probes).

The coding sequences of *FLM- β* , *FLM- δ* and *SVP* were amplified from cDNA and cloned into the pSPUTK expression vector. All generated constructs were confirmed by sequencing, and proteins were synthesized using TNT SP6 Quick Coupled Transcription/Translation System (Promega) according to the instructions of the manufacturer. To establish protein–protein interactions, protein synthesis was done in a single tube reaction for each protein complex. To ensure equal transcription/translation efficiency for each reaction in the titration experiments, total input of plasmid was kept equal by addition of pSPUTK expression vector containing a synthetically produced gene (GeneScript). This gene was designed by shuffling the *SVP* codon sequence, keeping the start codon at the first position and adding a premature stop codon after 456 bp to be able to distinguish the protein from *SVP* and *FLM* isoforms based on size.

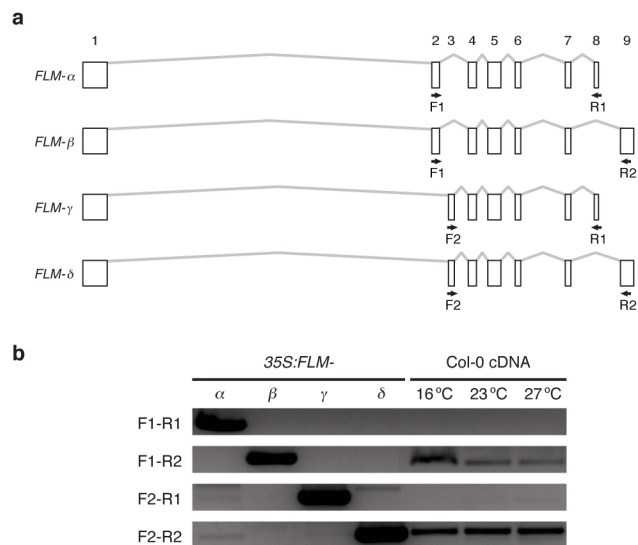
The binding reaction was performed in a reaction mixture containing 1.2 mM EDTA, pH 8.0, 0.25 mg ml^{−1} BSA, 7.2 mM HEPES, pH 7.3, 0.7 mM dithiothreitol, 60 μ g ml^{−1} salmon sperm DNA, 1.3 mM spermidine, 2.5% HAPS, 8% glycerol, 3.3 nmol ml^{−1} double-labelled double-stranded DNA (in case of *SEP3*) or 6.6 nm ml^{−1} single-labelled double-stranded DNA (in case of *SOC1* and *ATC*), and 2 μ l of *in-vitro*-synthesized proteins. Binding reaction was performed on ice for 45 min and loaded on a 5% polyacrylamide TBE gel. The gel was run at room temperature, followed by 2 h blotting to nylon membrane (Hybond-N+; Amersham, GE Healthcare Life Sciences). DNA shift was detected using the Chemiluminescent Nucleic Acid Detection Module (Pierce) according to the instructions of the manufacturer.

SEP3 probe (pGEM-T sequence underlined): 5′-CATGGCCGCGGGATTTGACGATAACTCCATCTTTCTATTTTGGGTAACGAGGTCCCTTCCCATTACGTCTTGACGTGGACCCTGTCCGCTCTATTTTACGAGAATCACTAGTGC GGCCGC-3′; *SOC1* probe: 5′-CGCTTGAAACCTCATCCTTACTTATTTTGAAAAAGCCTTAAGAAAGACCAAAATAGCATATTTTGATACATATG GACATTTTACATACATC-3′; *ATC* probe: 5′-TGGGTGCCCAACATTAA CATTTCACAAAAATGGTAAGTCCCAAGAATAATATTAGTTGTTTGGGAT ATATTCTTTGCAATACATCC-3′; shuffled *SVP*: 5′-ATGGAAGATAAATCG

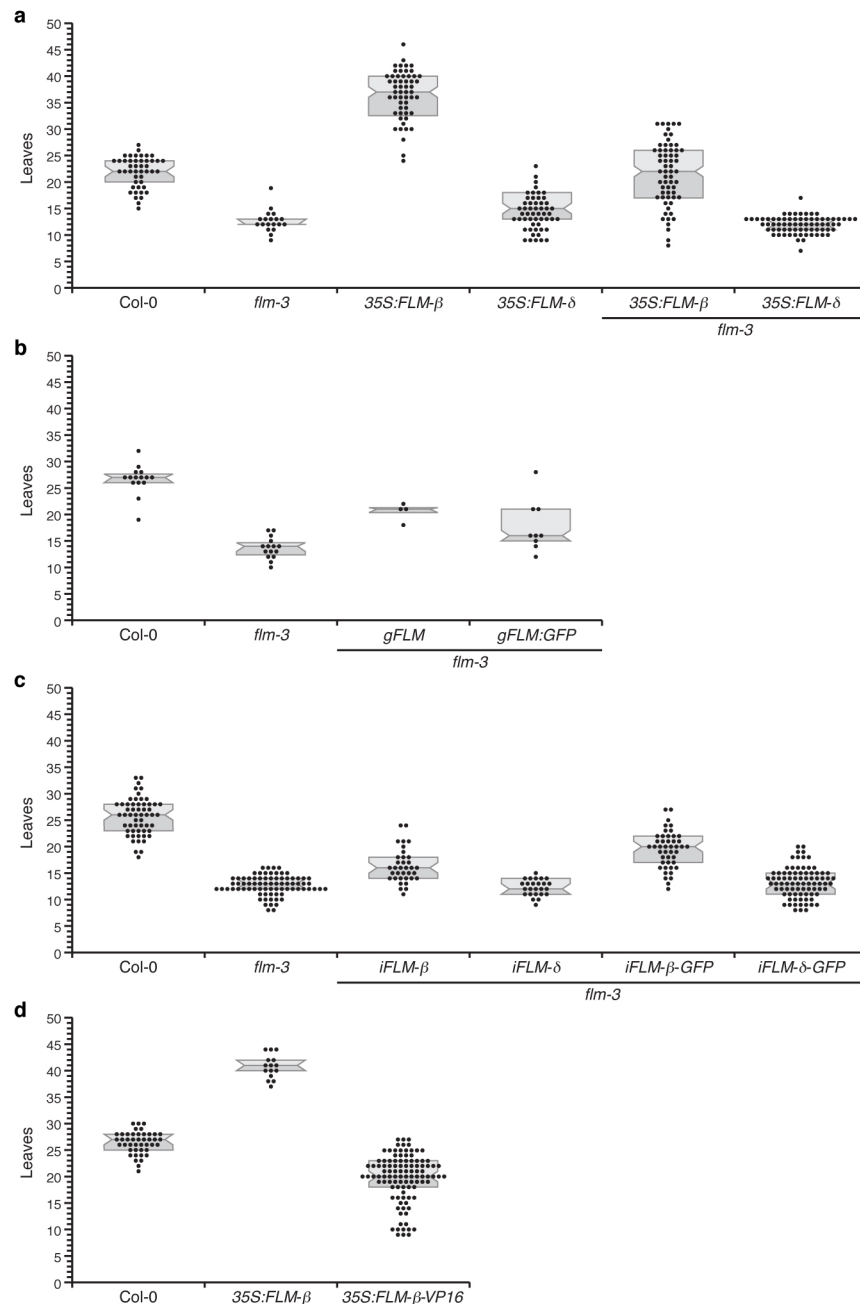
AAGTGTGCGGACACTAAGGAATTCGAAACGAGGATCGCGGTTGAAAT CGCCGCGCGCAATCCGAATACTCCTCCAGTGCTATTGAGGCTCTCCA TCAAAAGGAGACTAACAAGATGGGACTTAGCGTCTTGGCGCACCAGA TCTCGGAATCTGAAAGCAGAAAGTCGCACCAAGAAATTTAGCACTTC TTAAGATGATTCTAGGATTCCAGCCTAAGATCATGAGCCTTCTATGCA AAGAGGTTAAAGACAGACAGCTTCAAGTTAAGCAAGGAACCTTCGGTG GAGGACACGGAAATCAACAAGCTATCTCGAAGGGGACGTGAGAGACT GCGATGTCCACATCTTAGGGCTGGAGAGAATAAAGATGCTGAACAGC TCACCGAAGGTGGTTCCGTGTTGACCGGAGAGATTGTCTTATTAAGA ACCGATTTGAAAGGTAGGAGATGTCTCTCTCAGCGCCAACAGTACAG AGAACCAGCTAGAACTTGATGGCAACATTATGCGACAGAGACACTCC TATGAGTCATTAACACAGAGGCCGACGAACCTTGGTCTCGGTGCCGGT CAAGGGGATGACGTGCAGAACGCATCCAAAGGAGGAGGACAAACGGATT CTTGAGAGAGTCTCTTGTGTTGGAGACGAGTAGCAGTAACATGAAGC TGGACCAGTTGATAGGAGACGAGGCTGAGATGCAAACGCCCGGC-3′.

BiFC analyses. The binary vectors were co-infiltrated into 3-week-old tobacco (*Nicotiana benthamiana*) leaves as previously described³⁴. The absorbance (A) at 600 nm of the *Agrobacterium* cultures transformed with the split-mCitrine-tagged FLM and SVP constructs was adjusted to 0.25 (Fig. 2b and Extended Data Fig. 8d, e) and 0.3 (Extended Data Fig. 4) before the infiltration. To reduce false positive interactions, competition assays were performed by adding increasing concentrations of untagged version of FLM (*pJM356*, 35S:*FLM- β* ; *pLY225*, 35S:*FLM- δ*) or SVP (*pDP58*, 35S:*SVP*) to the infiltration. For the nuclear marker, 35S:*NLS-mCherry* and the silencing suppressor 35S:*p19*, an A_{600 nm} of 0.1 was used.

25. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462–469 (2003).
26. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
27. Immink, R. G. *et al.* Characterization of *SOC1*'s central role in flowering by the identification of its upstream and downstream regulators. *Plant Physiol.* **160**, 433–449 (2012).
28. Yant, L. *et al.* Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor APETALA2. *Plant Cell* **22**, 2156–2170 (2010).
29. Moyroud, E. *et al.* Prediction of Regulatory Interactions from Genome Sequences Using a Biophysical Model for the *Arabidopsis* LEAFY Transcription Factor. *Plant Cell* **23**, 1293–1306 (2011).
30. James, P., Halladay, J. & Craig, E. A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425–1436 (1996).
31. de Folter, S. *et al.* Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *Plant Cell* **17**, 1424–1433 (2005).
32. Severing, E. I. *et al.* Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS ONE* **7**, e30524 (2012).
33. van Dijk, A. D. *et al.* Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein–protein interaction. *PLoS Comput. Biol.* **6**, e1001017 (2010).
34. de Felippes, F. F. & Weigel, D. Transient assays for the analysis of miRNA processing and function. *Methods Mol. Biol.* **592**, 255–264 (2010).
35. Masiero, S. *et al.* INCOMPOSITA: a MADS-box gene controlling prophyl development and floral meristem identity in *Antirrhinum*. *Development* **131**, 5981–5990 (2004).

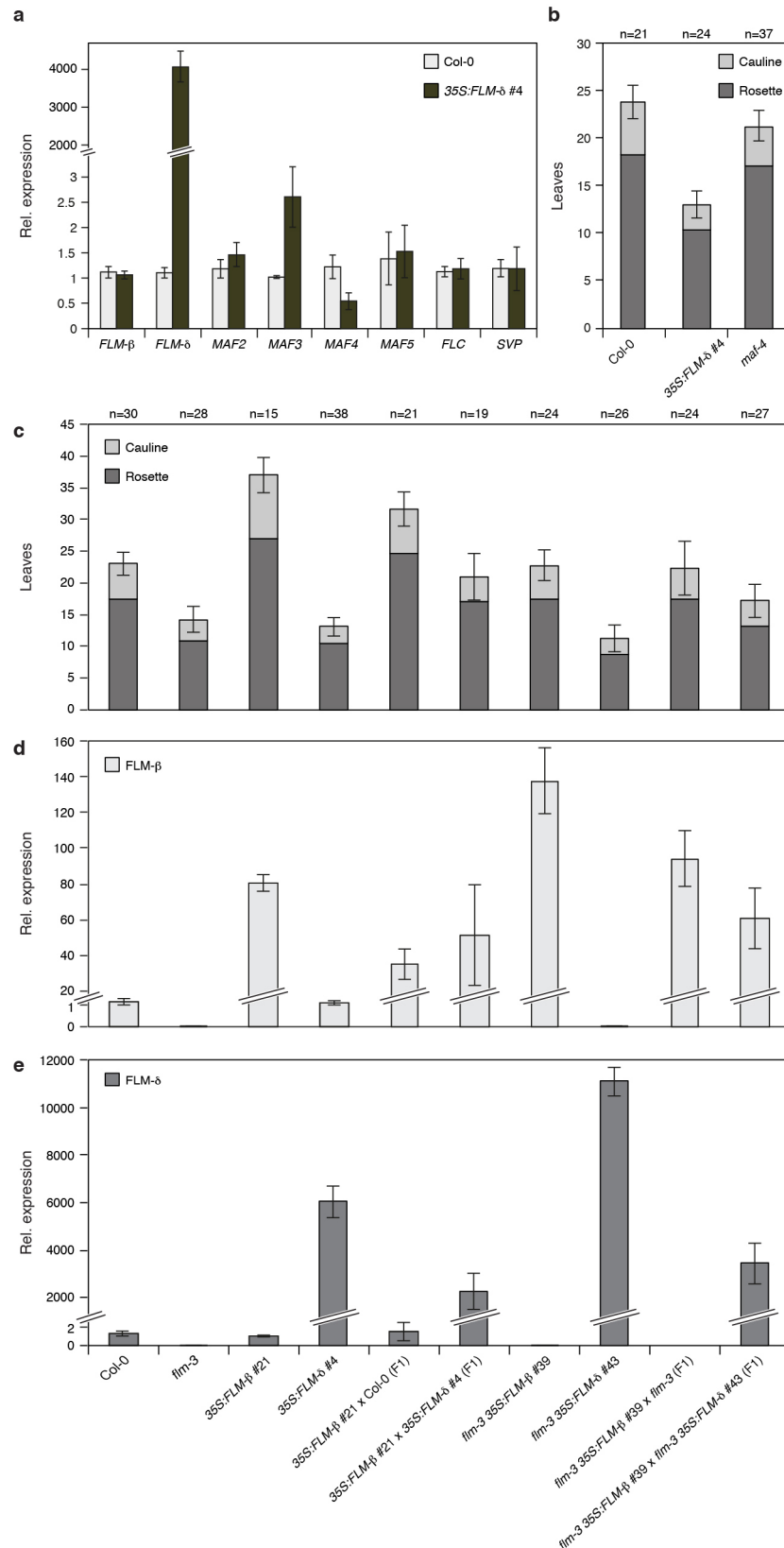


Extended Data Figure 1 | Analysis of *FLM* splice variant expression in Col-0. **a**, Graphic representation of the *FLM*- α , *FLM*- β , *FLM*- γ and *FLM*- δ transcripts, including exons (boxes) and introns (lines). Primers used for *FLM*- α (F1-R1), *FLM*- β (F1-R2), *FLM*- γ (F2-R1) and *FLM*- δ (F2-R2) amplification are shown. **b**, Semi-quantitative RT-PCR of *FLM* splice variants in Col-0 cDNA at different temperatures, using plasmids for each splice variant as controls (lanes 1–4).



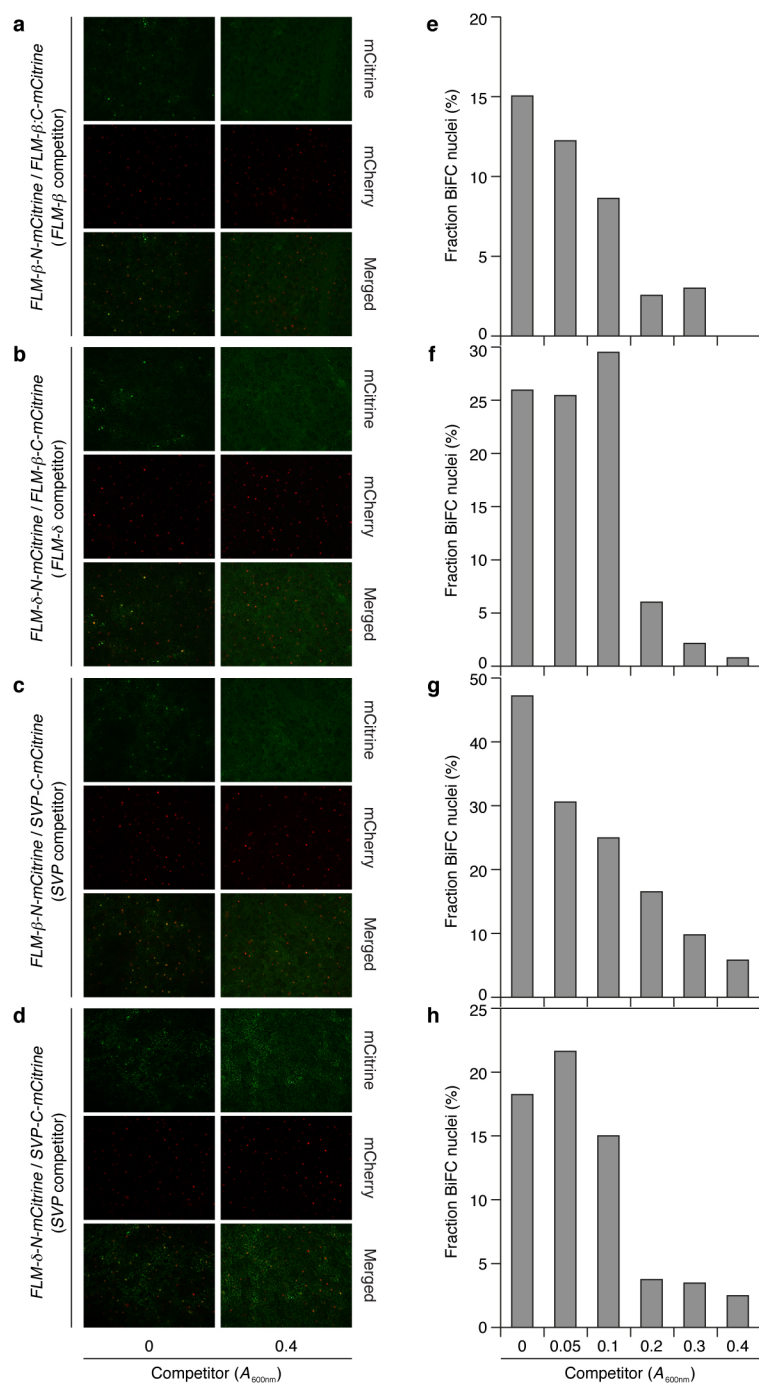
Extended Data Figure 2 | Distribution of flowering time of independent transgenic T1 lines established in this study. a–d, Flowering time of 35S:FLM- β and 35S:FLM- δ in Col-0 and *flm-3* mutant background (a), *flm-3* pFLM:gFLM and *flm-3* pFLM:gFLM-GFP (b), *flm-3* pFLM:iFLM- β , *flm-3*

pFLM:iFLM- δ , *flm-3* pFLM:iFLM- β -GFP and *flm-3* pFLM:iFLM- δ -GFP (c) and 35S:FLM- β -VP16 (d), grown under 16 °C, long-day are shown. Shaded areas mark the median and the 25% and 75% percentile of flowering time for a given genotype.



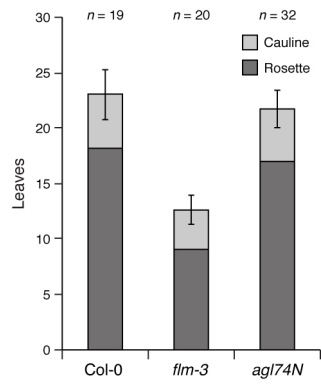
Extended Data Figure 3 | Analysis of *FLM* co-suppression in *35S:FLM- δ* .
a, Analysis of *FLM- β* , *FLM- δ* , *MAF2-5*, *FLC* and *SVP* in Col-0 control and *35S:FLM- δ* #4. All genes except *FLM- δ* and *MAF4* were expressed at similar levels in Col-0 and *FLM- δ* overexpression line. **b**, Flowering time of *maf-4* (SALK_028506) is similar to that of Col-0 control plants, indicating that the *MAF4* downregulation observed in **a** cannot explain the early flowering phenotype of the *35S:FLM- δ* line. **c–e**, Flowering time (**c**) and expression (**d**) of

FLM- β , and expression of *FLM- δ* (**e**), as determined by qRT-PCR analysis in F₁ populations from crosses between *35S:FLM- β* and *35S:FLM- δ* plants in both Col-0 and *flm-3* backgrounds. **d**, *FLM- β* expression is not co-suppressed in response to the *FLM- δ* misexpression (**e**) in both Col-0 and *flm-3* backgrounds. Rosette and cauline leaf numbers after bolting are represented in dark and light grey, respectively, in **b** and **c**. Error bars denote the s.d. of three biological replicates with three technical repetitions each in **a**, **d** and **e**.

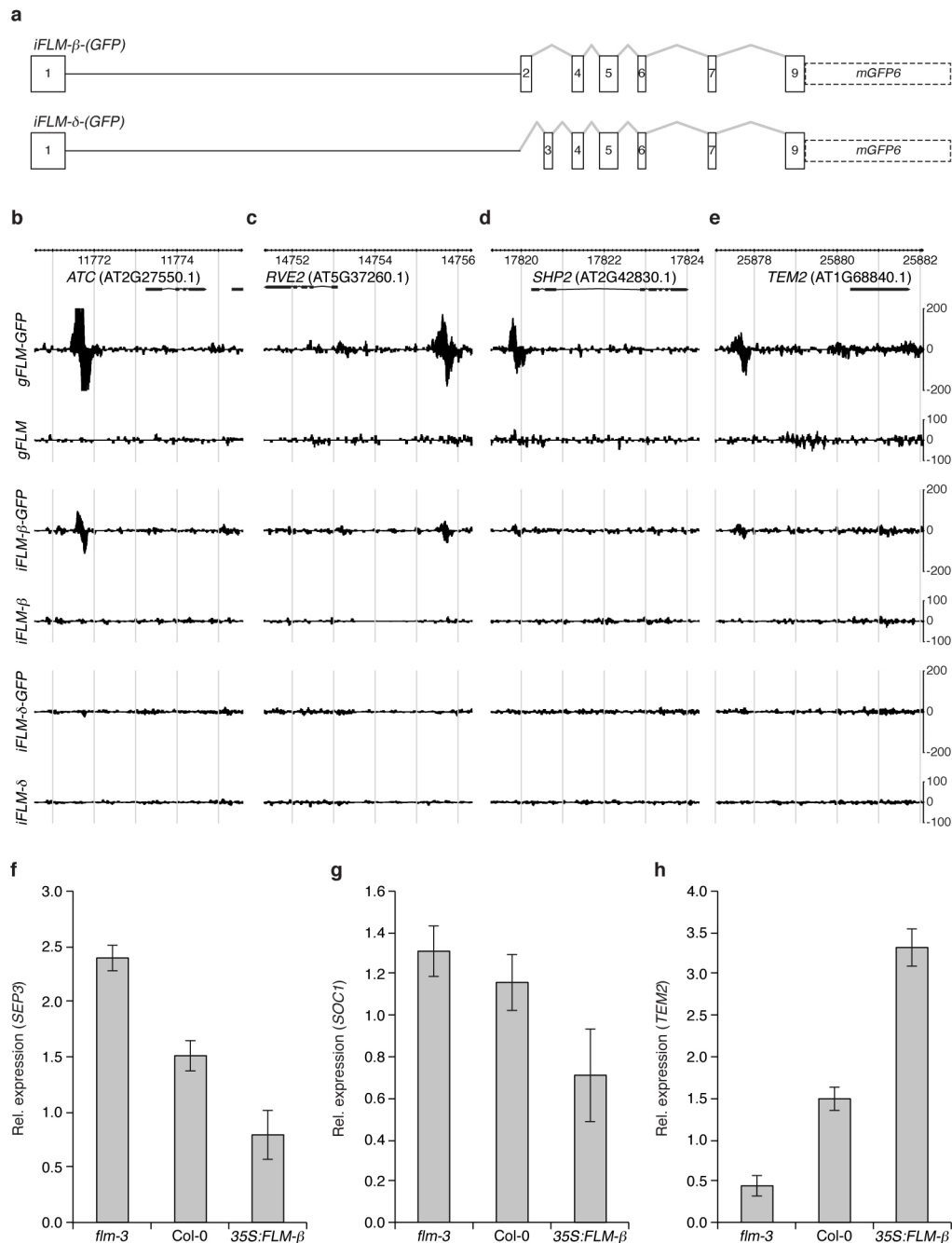


Extended Data Figure 4 | BiFC competition experiment. **a–d**, Microscope images (**a–d**) and quantification of mCitrine-positive nuclei (**e–h**). Increasing amounts ($A_{600\text{ nm}}$; bottom) of an untagged version of one of the interactors tested were included in the assay. The number of BiFC-positive nuclei decreases

with increasing amounts of the specific competitor: FLM- β –FLM- β homodimerization (**a**, **e**) and FLM- δ –FLM- β (**b**, **f**), SVP–FLM- β (**c**, **g**) and FLM- δ –SVP (**d**, **h**) heterodimerization.

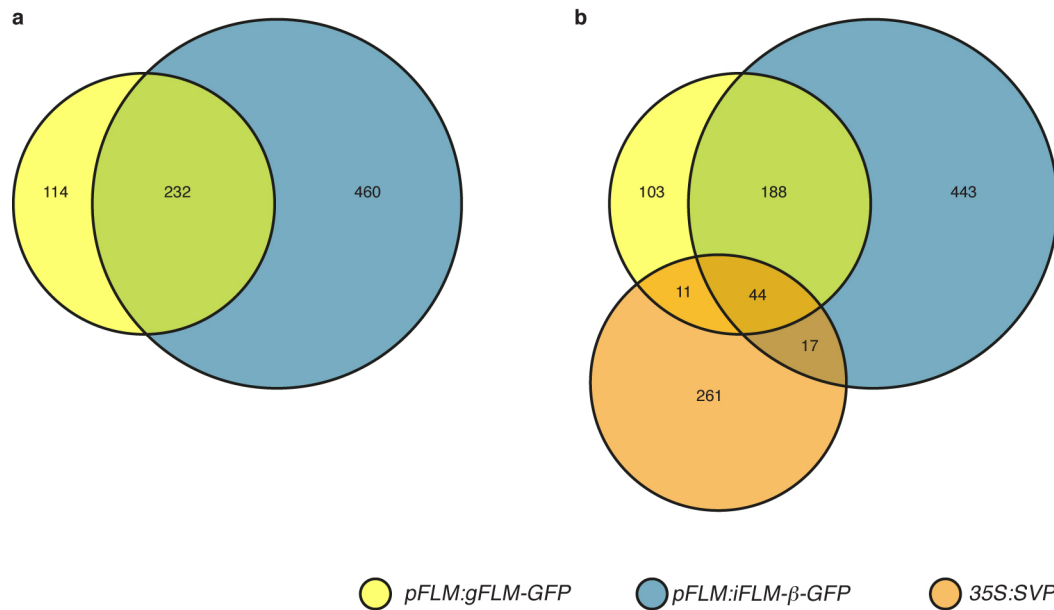


Extended Data Figure 5 | Flowering time of the *agl74N* T-DNA mutant.
 Flowering time of Col-0, *flm-3*, and a homozygous *agl74N* T-DNA insertion line (SALK_016446).



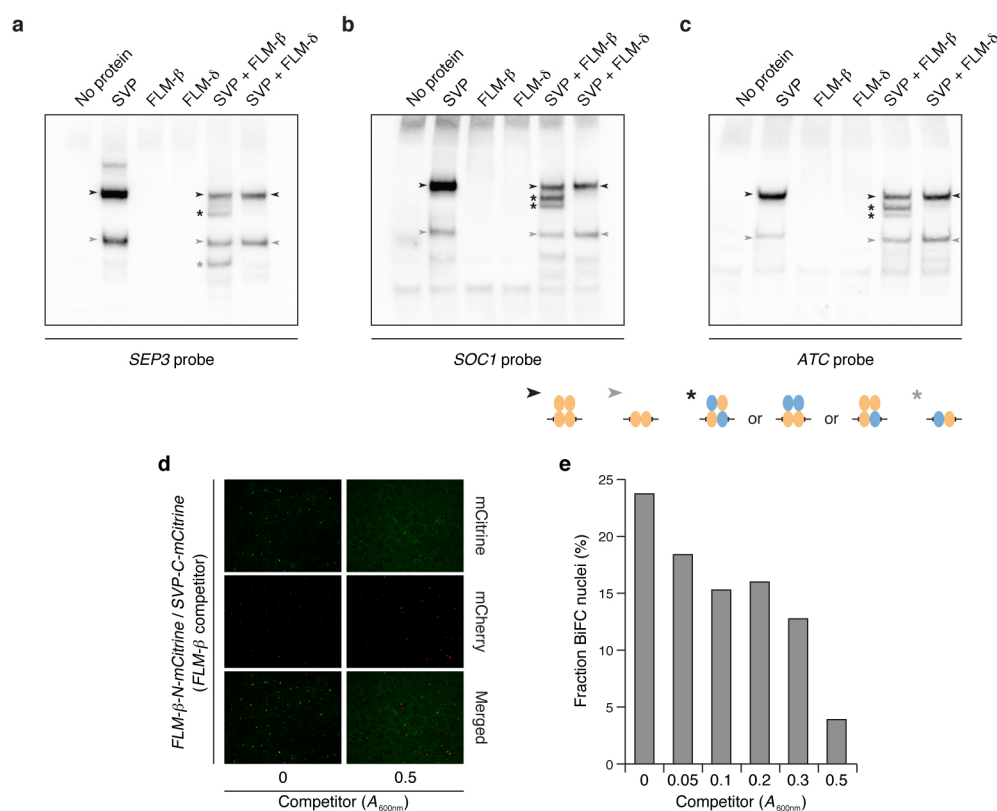
Extended Data Figure 6 | Graphic representation of *iFLM-β/δ-(GFP)* constructs, gBrowse traces of mapped ChIP-seq reads and validation of FLM targets. **a**, *iFLM-β/δ-(GFP)* constructs representation including exons (boxes), introns included (black flat line) and introns missing (grey lines). Dashed boxes indicate presence only in the *mGFP6*-tagged constructs. **b–e**, Local enrichment of FLM, *iFLM-β* and *iFLM-δ* binding in *ATC* (**b**), *RVE2* (**c**), *SHP2* (**d**) and *TEM2* (**e**). Chromosomal position (TAIR10) and models of the genes close to

the peaks are given at the top of the panels. Each panel shows a 5-kb window. Forward reads are mapped above each line and reverse reads below. **f–h**, qRT-PCR expression analysis of *SEP3* (**f**), *SOC1* (**g**) and *TEM2* (**h**) in *flm-3* mutant, Col-0 wild-type and a *35S:FLM-β* transgenic line show how increasing levels of FLM-β downregulate *SEP3* and *SOC1* expression, but induce *TEM2*. Error bars in **f–h** denote the s.d. of three biological replicates with three technical repetitions each.



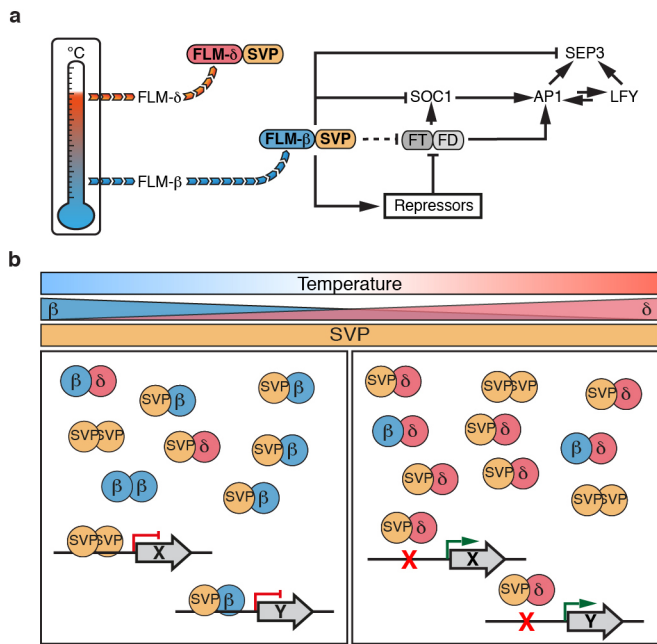
Extended Data Figure 7 | Venn diagram showing the number and overlap of FLM and SVP targets. **a**, Overlap of loci bound in *gFLM-GFP* and *iFLM-β-GFP* ChIP-seq experiments with a false discovery rate (FDR) < 0.1 in all biological replicates. At this FDR, the high quality *iFLM-β-GFP* data set identifies 460 targets that are missing from the *gFLM-GFP* data set, which

includes a replicate (#2) that contains substantial fewer uniquely mappable reads than the other replicates (see Supplementary Table 2). **b**, Overlap of loci bound in *gFLM-GFP* and *iFLM-β-GFP* ChIP-seq (FDR < 0.1) and SVP (FDR < 0.05) ChIP-chip assays²¹.



Extended Data Figure 8 | EMSA assays and FLM-β/SVP BiFC competition experiment. **a–c**, EMSA assay with three sequences identified as binding-sites for SVP²¹ and FLM (this work) by ChIP-chip and ChIP-seq, respectively. *SEP3* (**a**), *SOC1* (**b**) and *ATC* (**c**) promoter probes that include two (**a, b**) or one (**c**) CArG motif(s) were used in EMSA. Different order complexes are represented by black arrowheads and asterisks for homo- or heterotetramers, respectively, and with grey arrowheads and asterisks for homo- or

heterodimers, respectively. Orange and blue ellipses represent SVP and FLM-β proteins, respectively. **d, e**, Microscope images (**d**) and quantification (**e**) of mCitrine-positive nuclei. Increasing amounts (A_{600nm} ; bottom) of untagged 35S:FLM-δ were added to FLM-β and SVP mCitrine-tagged vectors. A reduction in the number of BiFC-positive nuclei is observed with increasing amounts of competitor.



Extended Data Figure 9 | Models of SVP-FLM complex function.

a, Temperature-dependent FLM splicing and genetic interactions of SVP-FLM- β heterocomplex in the flowering pathway. Strong binding of FLM to *FT* was observed in only one ChIP-seq replicate. Hence we propose that FLM-SVP downregulates *FT* expression in leaves indirectly through the induction of floral repressors transcription factors such as *TEM2* and the AP2-like *TOE3*. The FLM-SVP complex contributes to the repression of floral transition by directly downregulating *SOC1* and *SEP3* expression, where *SOC1* is a major floral activator. Arrows and block lines denote activation and repression, respectively. Dotted lines indicate a putative direct regulation. Rounded rectangles indicate proteins. **b**, Model of the temperature-dependent SVP-FLM complex function. Although *SVP* expression level is constant, *FLM- β* and *FLM- δ* levels are regulated in an antagonistic manner, with the former being the prevalent protein at low temperature and the latter dominating at high temperatures. At low temperatures SVP and FLM- β can interact, forming both homo- or heterocomplexes. The SVP-containing complexes are able to bind to the CArG boxes in the *cis* elements of important flowering related genes such as *SEP3*, *SOC1*, *ATC*, *TEM2* and *TOE3* and regulate their expression. When temperature increases, alternative splicing of *FLM* occurs, making *FLM- δ* the predominant splice variant. FLM- δ proteins compete with the remaining FLM- β and SVP proteins for complex formation. This results in the formation of non-functional SVP-FLM- δ complexes, which are impaired in their DNA-binding capability. The temperature-dependent splicing regulation of *FLM* occurs within 24 h, allowing the plant to quickly sense and respond to changes in ambient temperature, ensuring the switch between the non-flowering and flowering phase of development.

Extended Data Table 1 | Mutants and transgenic lines used in this study.

Line	Source
<i>flm-3</i>	(ref. 3)
<i>svp-32</i>	(ref. 4)
<i>svp-32/flm-3</i>	this study
<i>maf4</i>	SALK_028506, B. Davies
<i>agl74N</i>	this study; SALK_016446
<i>35S:FLM-β</i>	this study
<i>35S:FLM-δ</i>	this study
<i>flm-3 35S:FLM-β</i>	this study
<i>flm-3 35S:FLM-δ</i>	this study
<i>35S:SVP</i>	P. Huijser (ref. 35)
<i>svp-32 35S:FLM-β</i>	this study
<i>flm-3 35S:SVP</i>	this study
<i>flm-3 pFLM:gFLM</i>	this study
<i>flm-3 pFLM:gFLM-GFP</i>	this study
<i>flm-3 pFLM:FLM-β</i>	this study
<i>flm-3 pFLM:FLM-δ</i>	this study
<i>flm-3 pFLM:iFLM-β</i>	this study
<i>flm-3 pFLM:iFLM-δ</i>	this study
<i>flm-3 pFLM:iFLM-β-GFP</i>	this study
<i>flm-3 pFLM:iFLM-δ-GFP</i>	this study
<i>35S:FLM-β-VP16</i>	this study

Precision is essential for efficient catalysis in an evolved Kemp eliminase

Rebecca Blomberg^{1†}, Hajo Kries¹, Daniel M. Pinkas¹, Peer R. E. Mittl², Markus G. Grütter², Heidi K. Privett^{3†}, Stephen L. Mayo⁴ & Donald Hilvert¹

Linus Pauling established the conceptual framework for understanding and mimicking enzymes more than six decades ago¹. The notion that enzymes selectively stabilize the rate-limiting transition state of the catalysed reaction relative to the bound ground state reduces the problem of design to one of molecular recognition. Nevertheless, past attempts to capitalize on this idea, for example by using transition state analogues to elicit antibodies with catalytic activities², have generally failed to deliver true enzymatic rates. The advent of computational design approaches, combined with directed evolution, has provided an opportunity to revisit this problem. Starting from a computationally designed catalyst for the Kemp elimination³—a well-studied model system for proton transfer from carbon—we show that an artificial enzyme can be evolved that accelerates an elementary chemical reaction 6×10^8 -fold, approaching the exceptional efficiency of highly optimized natural enzymes such as triosephosphate isomerase. A 1.09 Å resolution crystal structure of the evolved enzyme indicates that familiar catalytic strategies such as shape complementarity and precisely placed catalytic groups can be successfully harnessed to afford such high rate accelerations, making us optimistic about the prospects of designing more sophisticated catalysts.

Proton transfer from carbon enables an impressive array of racemizations, carboxylations, eliminations, isomerizations and carbon–carbon bond-forming reactions. Although such reactions often have considerable kinetic and thermodynamic barriers, the enzymes that have evolved to catalyse them are among the most efficient known. For example, triosephosphate isomerase (TIM) accelerates the conversion of dihydroxyacetone phosphate to *R*-glyceraldehyde-3-phosphate, a key step in glycolysis, more than a billion-fold relative to the spontaneous reaction at neutral pH⁴. Because its efficiency is limited by diffusive steps rather than chemistry⁵, TIM is considered a ‘perfect’ enzyme.

Detailed biochemical and structural studies, coupled with mutagenesis and mechanistic data, have provided good qualitative understanding of TIM catalysis^{6,7}. The substrate binds in a shape complementary pocket, sequestered from bulk solvent by the closing of a mobile protein loop, where it is deprotonated by the carboxylate side chain of Glu 165. The side chains of His 95 and Lys 12 facilitate this process by electrostatically stabilizing the resulting enolate intermediate and flanking transition states. Reprotonation of the intermediate by Glu 165 yields *R*-glyceraldehyde-3-phosphate, which is released from the enzyme in a slow step to complete the catalytic cycle. Accessory interactions at the active site accurately place the catalytic residues and fine-tune their reactivity. Binding interactions with the non-reacting phosphodianion of the substrate further contribute to catalytic efficiency by positioning the substrate within the active site and activating the catalytic residues⁸.

In attempts to recapitulate such features in artificial catalysts, the base-promoted deprotonation of 5-nitrobenzisoxazole (**1**, Fig. 1)—the so-called Kemp elimination⁹—has been extensively studied. This abiological reaction is roughly 30 times faster than deprotonation of dihydroxyacetone

phosphate^{9,10}, but experimentally more tractable because it occurs by a one-step mechanism in which proton transfer is coupled to irreversible cleavage of the N–O bond. Although many catalysts for the Kemp elimination have been described, including catalytic antibodies¹¹ and computationally designed enzymes^{12,13}, they typically effect proton abstraction with relatively modest turnover numbers ($k_{\text{cat}} < 1 \text{ s}^{-1}$). Even after extensive directed evolution, the most active artificial enzyme generated so far, KE59.13 (ref. 14), cleaves 5-nitrobenzisoxazole with a k_{cat} of only 9.5 s^{-1} and an apparent second-order rate constant ($k_{\text{cat}}/K_{\text{m}}$) of $60,000 \text{ M}^{-1} \text{ s}^{-1}$. Like TIM, it uses a glutamate side chain as its catalytic base, albeit much less efficiently.

Here we show that effective placement of appropriate catalytic groups in the right active site environment can bridge the efficiency gap between natural and artificial enzymes. As our starting point, we used *in silico* design HG3 (ref. 3), which catalyses the Kemp elimination of 5-nitrobenzisoxazole with catalytic parameters $k_{\text{cat}} = 0.68 \text{ s}^{-1}$ and $k_{\text{cat}}/K_{\text{m}} = 430 \text{ M}^{-1} \text{ s}^{-1}$. Optimization of HG3 was based on an evolutionary strategy that included both global and local mutagenesis (Supplementary Fig. 1). Error-prone PCR and DNA shuffling were used to target the entire gene and identify ‘hot spots’ (rounds 1–14; 16–17); small focused libraries were subsequently constructed to interrogate these and other positions in the active site and substrate entry tunnel (rounds 1b; 15). On average, only one amino acid mutation per evolutionary round was fixed (Fig. 2a, b and Supplementary Fig. 2), improving the likelihood of only accumulating beneficial changes along the evolutionary trajectory. In contrast to many laboratory optimization experiments, catalytic improvements during the evolution of HG3 were accompanied by increased thermostability (change

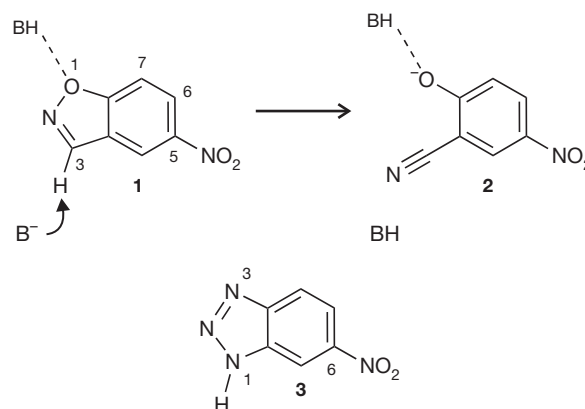


Figure 1 | Kemp elimination. Deprotonation of 5-nitrobenzisoxazole (**1**) affords a salicylonitrile (**2**). The transition state analogue 6-nitrobenzotriazole (**3**) has an acidic N–H bond in place of the scissile C–H in the substrate.

¹Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland. ²Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland. ³Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA. ⁴Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA. [†]Present addresses: Corporate R&D Division, Firmenich SA, 1211 Geneva, Switzerland (R.B.); Protabit, Pasadena, California 91101, USA (H.K.P.).

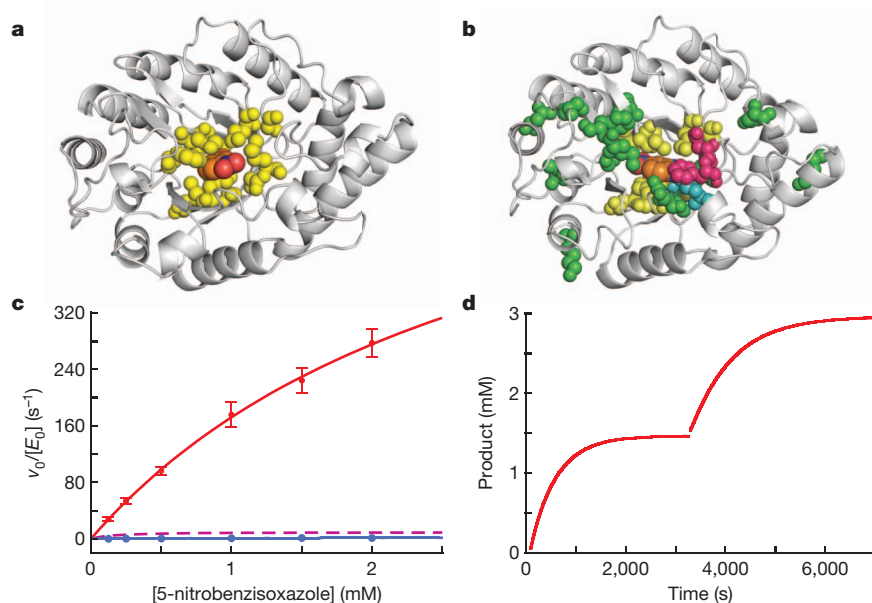


Figure 2 | Directed evolution of Kemp eliminase HG3. **a**, The computationally designed enzyme HG3 was generated by introducing 11 point mutations (yellow) into a xylanase scaffold. The modelled transition state for the Kemp elimination is shown with orange carbons. **b**, The HG3.17 variant was obtained after 17 rounds of directed evolution. Residues mutated once are shown in green, twice in cyan, and three times in magenta. **c**, Michaelis–Menten plots for HG3 (blue) and HG3.17 (red). The HG3.17 data represent the average of six separate protein measurements from three independent protein batches, with error bars denoting s.d. The analogous curve for the previously evolved Kemp eliminase KE59.13 (ref. 14) is shown as magenta dashes. **d**, Progress curve for consecutive conversion of two 1.5 mM batches of 5-nitrobenzisoxazole by 10 nM HG3.17.

in melting temperature (ΔT_M) = 7 °C; Supplementary Fig. 3a) and protein yield (~10-fold).

The most active variant to emerge after 17 rounds of mutagenesis and screening cleaves 5-nitrobenzisoxazole with a k_{cat} of $700 \pm 60 \text{ s}^{-1}$ (mean \pm s.d.) and a k_{cat}/K_m of $230,000 \pm 20,000 \text{ M}^{-1} \text{ s}^{-1}$ (Fig. 2c and Supplementary Table 1). For comparison, the k_{cat} and k_{cat}/K_m values for TIM-catalysed conversion of dihydroxyacetone phosphate to *D*-glyceraldehyde-3-phosphate are 430 s^{-1} and $440,000 \text{ M}^{-1} \text{ s}^{-1}$, respectively¹⁵. HG3.17 uses the carboxylate side chain of Asp 127 for proton abstraction ($pK_a = 6.0$, Supplementary Fig. 3b); replacement of this residue with alanine or asparagine leads to a $>10^5$ -fold reduction in activity (Supplementary Fig. 3c). The active site carboxylate is 4×10^9 -fold more effective than a simple carboxylate base such as acetate in aqueous solution (Supplementary Table 1). Notably, catalytic improvement over the original computational design was achieved entirely by increasing the turnover number, not substrate affinity. As a consequence, HG3.17 promotes proton transfer from 5-nitrobenzisoxazole with a 6×10^8 -fold rate acceleration over the uncatalysed reaction (k_{cat}/k_{uncat}), outperforming other Kemp eliminases by nearly two orders of magnitude^{11,14} (Fig. 2c). Given the importance of high turnover numbers for industrial biocatalysis, where rapid conversion of substrate to product is desired, such high reactivity is noteworthy. Product inhibition, a common problem in enzyme catalysis, is also negligible in the case of HG3.17, which readily achieves $>3 \times 10^5$ turnovers per active site (Fig. 2d).

These results demonstrate that a bottom-up approach combining *de novo* computational design with laboratory evolution can yield artificial biocatalysts capable of accelerating proton transfer with true enzymatic efficiency. To determine the origins of its catalytic prowess, the HG3.17 enzyme was co-crystallized with 6-nitrobenzotriazole (3). This stable transition state analogue, which inhibits the enzyme with an inhibition constant (K_i) of 2 μM (Supplementary Fig. 3d), has an acidic N–H bond ($pK_a = 6.3$) that mimics the scissile C–H bond of the substrate. The X-ray structure of the complex, solved at 1.09 Å resolution, shows the ligand deeply buried in the active site (Fig. 3a), shielded from bulk solvent and oriented in a catalytically productive pose (Fig. 3b). Comparison of HG3.17 with the starting HG3 enzyme³ suggests that three factors were decisive in the evolution of high activity.

First, the evolved active site exhibits extraordinarily high shape complementarity to the ligand. After binding, approximately 95% of the ligand surface is buried (Fig. 3a). A tight fit is achieved by a combination of side chain and backbone interactions plus a single ordered water

molecule. Recognition of small apolar ligands such as 5-nitrobenzisoxazole and 6-nitrobenzotriazole constitutes a challenge for computational design, and ambiguous ligand binding modes are not uncommon¹⁴. This was the case with the computationally designed HG3 enzyme, which binds 6-nitrobenzotriazole in two different orientations³. By contrast, HG3.17 displays only a single binding mode in which the ligand is flipped relative to the original design model³. Excluding non-productive binding modes in this way (Fig. 4a) represents a relatively simple means of increasing catalytic efficiency. In addition to aiding catalysis by maximizing attractive interactions with the transition state¹⁶, snug binding can enhance substrate specificity¹⁷. In contrast to many Kemp eliminases that accept a wide range of substituted benzisoxazoles^{14,18}, HG3.17 is rather discriminating. Benzisoxazoles lacking a substituent at the 5 position or containing groups at the 6 or 7 positions are much poorer substrates than intrinsic reactivity would predict (Supplementary Fig. 3e).

Second, ligand alignment with Asp 127, the catalytic base, was optimized by evolution. Asp 127 sits at the bottom of a hydrophobic binding pocket (Fig. 3a). One carboxylate oxygen is hydrogen bonded to a buried water molecule, whereas the other forms a tight hydrogen bond with the acidic proton of the bound benzotriazole (Fig. 3b). The hydrogen bond to the ligand is unusually short (2.53 Å)—significantly shorter than in other Kemp eliminases^{14,19}—and its geometry converged towards optimal values for hydrogen bonding interactions over the evolutionary trajectory (Fig. 4b and Supplementary Fig. 4). Notably, although the syn orbital of a carboxylate has been argued to be more basic than the anti orbital²⁰, this interaction involves the anti orbital of the Asp 127 carboxylate (Fig. 3b). The efficiency of HG3.17 suggests that the energy penalty of using the ‘wrong’ lone pair for proton abstraction may be small if the base and substrate are precisely aligned. A nonspecific medium effect does not seem to be the source of the observed boost in catalytic efficiency. Although the apolar environment of the active site undoubtedly contributes to the high reactivity of the desolvated carboxylate ion¹⁸, the starting design and the evolved catalyst have similar pH profiles (Supplementary Fig. 3b), arguing against a substantial change in either Asp 127 basicity or the desolvation penalty associated with substrate binding.

Third and perhaps most importantly, a new catalytic group capable of stabilizing developing negative charge in the transition state emerged over the course of evolution (Fig. 4c). This innovation, which is reminiscent of the oxyanion hole in serine proteases²¹, sets HG3.17 apart from other Kemp eliminases^{11,14}. Although Thr 265 was originally supposed

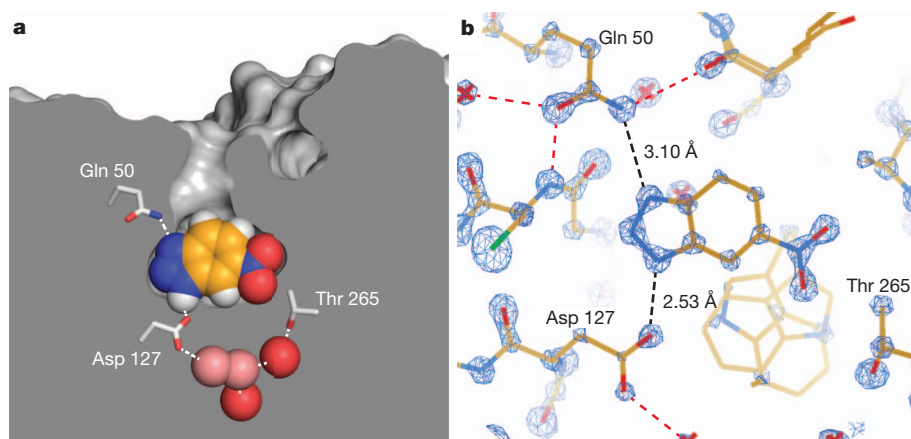


Figure 3 | Crystal structure of HG3.17 complexed with 6-nitrobenzotriazole. **a**, Cut-away view of the active site showing the snug fit of the ligand within the binding pocket. The ligand is shown in space-filling representation (carbon, orange; nitrogen, blue; oxygen, red; hydrogen, white). Critical residues are shown as sticks (carbon, white). Ordered water molecules in the active site are shown as spheres (red, full occupancy; salmon, partial occupancy). **b**, Electron density for the inhibitor interacting with Asp 127 and Gln 50. The $2F_o - F_c$ map is contoured at 5.5σ . Dashed lines indicate hydrogen bonds.

to donate a hydrogen bond to the phenoxide leaving group³, the flipped orientation of the substrate in the evolved variant precludes such an interaction. Instead, the side chain of Gln 50 assumes this role, making a direct hydrogen bonding contact (3.10 Å) with N3 of 6-nitrobenzotriazole (Fig. 3b). This residue, which was mutated twice during evolution, first from lysine to histidine and later from histidine to glutamine (Supplementary Fig. 2), is well placed to position the substrate for proton abstraction and simultaneously facilitate charge transfer from the buried carboxylate to the phenolic oxygen. Consistent with this hypothesis, replacement of Gln 50 by alanine reduces the k_{cat}/K_m by 50-fold (Supplementary Fig. 3c). Two supporting hydrogen-bonding interactions with backbone amides anchor the side-chain amide in a catalytically productive orientation. Complementing the significant entropic advantage expected for effectively positioning this functional group, the low dielectric environment of the binding pocket would be expected to enhance its interaction with the charged transition state²².

Creation of highly preorganized environments sensitive to the small changes in structure and charge distribution that distinguish ground and transition state is extraordinarily challenging²³. As a consequence, reducing the activation energy of an elementary chemical step is considered to be the most demanding aspect of enzyme evolution²⁴. That effective transition-state tuning could be achieved for the Kemp elimination through a combination of computational design and experimental evolution is thus good news for enzyme engineers. One may nevertheless wonder why computation alone was not more successful, particularly as the evolved HG3.17 resembles the idealized active site targeted by design³. The answer probably lies in the tradeoffs between speed and accuracy that current design algorithms make to facilitate searches of sequence space²⁵. The discrete amino acid side-chain rotamers and ligand placement schemes do not allow the side-chain functional groups (or bound ligand) to adopt conformations with the sub-Ångström

precision necessary for high catalytic efficiency. Inaccurate treatment of long-range electrostatic interactions, of paramount importance in enzymatic catalysis²⁶, and the simplifying assumption that the protein scaffold is rigid compound this problem.

More advanced computational techniques will be needed to tackle these challenges. In this regard, ‘multistate’ design may have a role in lessening the severity of the fixed backbone assumption by allowing the sequence design calculation to sample from a large ensemble of closely related structural states²⁷. Using an ensemble of states could capture the backbone structural variations needed to allow for a more precise positioning of relevant catalytic side chains. In addition, classical molecular dynamics and combined quantum mechanics/molecular mechanics simulations have shown some promise in evaluating and ranking designs²⁸, and will probably become increasingly important for identifying improved lead candidates for experimental optimization.

Although HG3.17 catalyses the deprotonation of 5-nitrobenzisoxazole with remarkable efficiency, its apparent second-order rate constant k_{cat}/K_m is still three to four orders of magnitude below the diffusion limit (10^8 – 10^9 M⁻¹ s⁻¹). Future efforts to turn HG3.17 into a ‘perfect’ enzyme⁶ might profitably focus on optimizing the dynamic properties of the protein. Because 6-nitrobenzotriazole is almost completely buried within the binding pocket (Fig. 3a), ligand association and dissociation necessarily require considerable protein conformational changes. Although many beneficial mutations accumulated near the mouth of the binding pocket (Fig. 2b), shortening and widening the substrate entry tunnel relative to the starting design (Supplementary Fig. 3f), structural rearrangements may still hinder enzyme–substrate encounter¹⁷.

Design and evolution of a highly active Kemp eliminase demonstrates the feasibility of mimicking the precisely tailored active sites of true enzymes. Although the catalytic devices that emerged during HG3 optimization (Fig. 4) are familiar from top-down studies of numerous natural enzymes, their successful implementation required exactitude

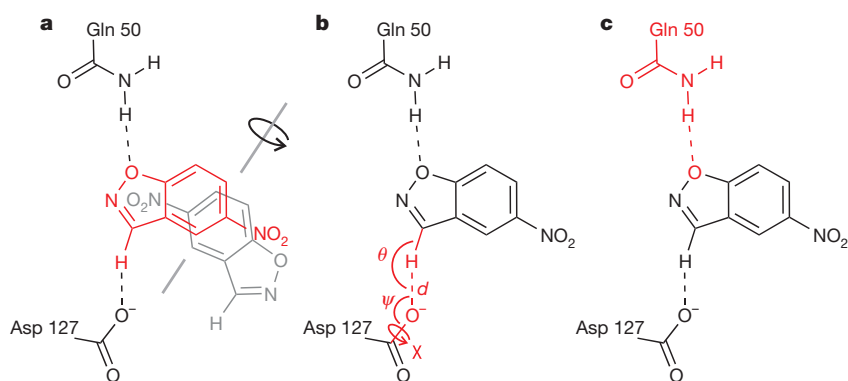


Figure 4 | Catalytic improvement of HG3. **a**, High shape complementarity between substrate and active site eliminated unproductive binding modes (grey molecule). **b**, Efficient proton transfer was facilitated by optimizing the interactions between the catalytic base and the bound ligand. **c**, Introduction of an oxyanion binder contributed to transition state stabilization.

beyond the current capabilities of computational methods. In the future, generalizing *de novo* enzyme design to other chemical reactions will undoubtedly profit from improved algorithms and increased computing power^{29,30}. In the quest for enzyme-like activities, however, it is likely that designed active sites will continue to require adjustment and refinement through evolution, nature's own optimization strategy.

METHODS SUMMARY

In vitro evolution. Gene libraries of HG3 were generated by error-prone PCR using the GeneMorph II Random Mutagenesis kit (Stratagene) and by DNA shuffling of the most active variants from each round. Focused libraries were generated by conventional site-directed mutagenesis using oligonucleotides containing degenerate codons. Kemp eliminase activity of 800–1,000 variants per round was assayed with 5-nitrobenzisoxazole in 96-well plates as previously described⁴. The clones with the largest increase in activity, typically corresponding to about 1% of the screened population, were picked from replica plates for plasmid isolation, sequencing and further diversification.

Biochemical characterization. All HG3 variants were produced as carboxy-terminally His-tagged proteins in *Escherichia coli* BL21-Gold(DE3), and purified by affinity (Ni-NTA) and cation exchange chromatography. Individual point mutants were generated by conventional site-directed mutagenesis. Cleavage of 5-nitrobenzisoxazole was assayed at 27 °C in 50 mM phosphate buffer, 100 mM NaCl containing 10% methanol, pH 7.0. Product formation was monitored spectroscopically at 380 nm ($\Delta\epsilon = 15,800 \text{ M}^{-1} \text{ cm}^{-1}$). Steady-state parameters were obtained by fitting the data to the Michaelis–Menten equation.

Crystallization and structure determination. A variant of HG3.17 containing reversions of two surface mutations (Asn47Glu/Asp300Asn) and complexed with transition state analogue 6-nitrobenzotriazole (**3**) was crystallized by vapour diffusion in sitting drops. Diffraction data were collected at 100 K at the Swiss Light Source (SLS) X06SA PX beamline. The structure was solved by molecular replacement with PHASER and refined using the programs PHENIX and COOT. Refinement statistics are summarized in Supplementary Table 2.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June; accepted 30 August 2013.

Published online 16 October 2013.

- Pauling, L. Molecular architecture and biological reactions. *Chem. Eng. News* **24**, 1375–1377 (1946).
- Hilvert, D. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751–793 (2000).
- Privett, H. K. *et al.* Iterative approach to computational enzyme design. *Proc. Natl Acad. Sci. USA* **109**, 3790–3795 (2012).
- Albery, W. J. & Knowles, J. R. Free-energy profile for the reaction catalyzed by triosephosphate isomerase. *Biochemistry* **15**, 5627–5631 (1976).
- Blacklow, S. C., Raines, R. T., Lim, W. A., Zamore, P. D. & Knowles, J. R. Triosephosphate isomerase catalysis is diffusion controlled. *Biochemistry* **27**, 1158–1165 (1988).
- Knowles, J. R. Enzyme catalysis: not different, just better. *Nature* **350**, 121–124 (1991).
- Wierenga, R. K., Kapetaniou, E. G. & Venkatesan, R. Triosephosphate isomerase: A highly evolved biocatalyst. *Cell. Mol. Life Sci.* **67**, 3961–3982 (2010).
- Richard, J. P. A paradigm for enzyme-catalyzed proton transfer at carbon: Triosephosphate isomerase. *Biochemistry* **51**, 2652–2661 (2012).
- Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. Physical organic-chemistry of benzisoxazoles. 1. Mechanism of base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38**, 2294–2301 (1973).
- Richard, J. P. Acid-base catalysis of the elimination and isomerization-reactions of triose phosphates. *J. Am. Chem. Soc.* **106**, 4926–4936 (1984).
- Thorn, S. N., Daniels, R. G., Auditor, M. T. M. & Hilvert, D. Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **373**, 228–230 (1995).

- Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
- Korendovych, I. V. *et al.* Design of a switchable eliminase. *Proc. Natl Acad. Sci. USA* **108**, 6823–6827 (2011).
- Khersonsky, O. *et al.* Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl Acad. Sci. USA* **109**, 10358–10363 (2012).
- Putman, S. J., Coulson, A. F. W., Farley, I. R. T., Riddleston, B. & Knowles, J. R. Specificity and kinetics of triose phosphate isomerase from chicken muscle. *Biochem. J.* **129**, 301–310 (1972).
- Wolfenden, R. Enzyme catalysis: Conflicting requirements of substrate access and transition-state affinity. *Mol. Cell. Biochem.* **3**, 207–211 (1974).
- Herschlag, D. The role of induced fit and conformational changes of enzymes in specificity and catalysis. *Bioorg. Chem.* **16**, 62–96 (1988).
- Hu, Y., Houk, K. N., Kikuchi, K., Hotta, K. & Hilvert, D. Nonspecific medium effects versus specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *J. Am. Chem. Soc.* **126**, 8197–8205 (2004).
- Debler, E. W. *et al.* Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc. Natl Acad. Sci. USA* **102**, 4984–4989 (2005).
- Gandour, R. D. On the importance of orientation in general base catalysis by carboxylate. *Bioorg. Chem.* **10**, 169–176 (1981).
- Bryan, P., Pantoliano, M. W., Quill, S. G., Hsiao, H. Y. & Poulos, T. Site-directed mutagenesis and the role of the oxyanion hole in subtilisin. *Proc. Natl Acad. Sci. USA* **83**, 3743–3745 (1986).
- Na, J., Houk, K. N. & Hilvert, D. Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J. Am. Chem. Soc.* **118**, 6462–6471 (1996).
- Frushicheva, M. P., Cao, J., Chu, Z. T. & Warshel, A. Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proc. Natl Acad. Sci. USA* **107**, 16869–16874 (2010).
- Albery, W. J. & Knowles, J. R. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, 5631–5640 (1976).
- Voigt, C. A., Gordon, D. B. & Mayo, S. L. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **299**, 789–803 (2000).
- Warshel, A. *et al.* Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
- Allen, B. D., Nisthal, A. & Mayo, S. L. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proc. Natl Acad. Sci. USA* **107**, 19838–19843 (2010).
- Kiss, G., Röthlisberger, D., Baker, D. & Houk, K. N. Evaluation and ranking of enzyme designs. *Protein Sci.* **19**, 1760–1773 (2010).
- Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed. Engl.* **52**, 5700–5725 (2013).
- Kries, H., Blomberg, R. & Hilvert, D. *De novo* enzymes by computational design. *Curr. Opin. Chem. Biol.* **17**, 221–228 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors are grateful to A. Aires-Trapote and C. Mayer for experimental assistance. We also thank C. Stutz and B. Blattmann for help in protein crystallization, and the beamline staff at the Swiss Light Source for support during data collection. This work was supported by the Swiss National Science Foundation (SNSF), the National Center of Excellence in Research (NCCR) Structural Biology program of the SNSF, the ETH Zurich, and the Defense Advanced Research Projects Agency (DARPA). Fellowships from the Fonds des Verbandes der chemischen Industrie (to R.B.), the Stipendienfonds der Schweizer Chemischen Industrie (to H.K.), the Studienstiftung des deutschen Volkes (to R.B. and H.K.), and the National Security Science and Engineering Faculty Fellowship (to S.L.M.) are gratefully acknowledged.

Author Contributions D.H., M.G.G., S.L.M., H.K.P., P.R.E.M., D.M.P., H.K. and R.B. designed the experiments. R.B. and H.K. evolved and biochemically characterized the variants; D.M.P. and P.R.E.M. crystallized the proteins and solved their structures. The manuscript and figures were prepared by R.B., H.K., D.M.P. and D.H.

Author Information The crystal structure of HG3.17 was deposited in the RCSB Protein Data Bank (PDB) under the accession number 4BS0. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H. (hilvert@org.chem.ethz.ch).

METHODS

Starting point for *in vitro* evolution. Kemp eliminase HG3 (ref. 3) was designed into the thermostable xylanase TAX from *Thermoascus aurantiacus* (PDB accession code 1GOR). Eleven mutations were introduced into the inert scaffold (Gln42Met, Thr44Trp, Arg81Gly, His83Gly, Thr84Met, Asn130Gly, Asn172Met, Ala234Ser, Thr236Leu, Glu237Met and Trp267Phe) using the method described previously³¹. The resulting enzyme served as the starting point for several rounds of directed evolution via whole gene random mutagenesis, DNA shuffling, and the use of focused libraries.

Library construction. The gene encoding HG3 was subcloned into the commercial pET-11b(+) vector (Novagen). Primers for library cloning and the focused libraries (Supplementary Table 3) were purchased from Microsynth AG (Balgach). Diversity in gene libraries was created using error-prone PCR³², DNA shuffling³³ and combinatorial site-directed mutagenesis^{34,35}. Randomization experiments targeting the entire gene were carried out by error-prone PCR using the GeneMorph II Random Mutagenesis kit from Stratagene. Site-specific saturation mutagenesis targeting defined sites was achieved with primer sets that overlapped on the 5' side of the randomized position, together with appropriate flanking primers (T7 terminator and T7 alternative). Randomized codons were incorporated into the sense primer. Furthermore, a mismatch at the wobble position of the preceding codon was introduced to ensure unbiased incorporation of all possible base combinations. The length of the sense primers was determined by the desired melting temperature of the sequence ($T_M = 53 \pm 2^\circ\text{C}$) on the 3' and 5' sides of the mutation. Melting temperatures were calculated using the nearest neighbour method³⁶. The genes for all focused libraries and point mutants were generated by an overlap extension PCR procedure. First, gene fragments containing the desired mutations were produced by standard PCR using the primer combination described above and subsequently purified by agarose gel electrophoresis. Next, the gene fragments were assembled by combining 100 ng of each fragment for three PCR cycles, followed by addition of the flanking primers for amplification of the entire construct. The PCR products (inserts) were subsequently purified by agarose gel electrophoresis. Integration of the gene libraries into the library vectors pET-11b(+) was achieved by standard cloning techniques. In all cases, double restriction digests of the vectors and inserts were carried out using appropriate restriction endonucleases according to the manufacturer's instructions. The NdeI restriction endonuclease was used in fivefold excess and extended incubation times of at least 6 h to compensate for its weak activity. The vector and insert fragments were purified by agarose gel electrophoresis. The libraries were ligated at 16°C for 16–72 h with a 1:3 vector-to-insert ratio (final vector concentration 10 nM) and 30 U μl^{-1} T4 DNA ligase. Finally, the ligation mixtures were purified by phenol–chloroform extraction, desalted and concentrated in a Microcon YM-30 device (Millipore) before transformation of BL21-Gold(DE3) cells. The transformation mixture was plated on LB agar containing $150 \mu\text{g l}^{-1}$ ampicillin. Library quality and mutation rate were assessed by sequence analysis of plasmids from single colonies before screening.

Focused libraries round 1b. The sites for eight focused libraries were either based on the hot spots identified in the second round of random mutagenesis and screening (Val 6/Ile 10, Ser 89/Gln 90, Arg 124/Ala 125, Glu 132 and Met 172) or chosen by visual inspection of the published crystal structure (Lys 50/Met 84, Leu 236/Met 237 and Thr 265) (Supplementary Table 4a). DNA libraries were constructed by standard overlap extension PCR (as described above) using degenerate primers (Supplementary Table 3). Subsequently, beneficial mutations were assembled step-wise and the improvement factor was calculated from the activity rates determined in lysate screens (Supplementary Table 4b).

Focused libraries rounds 15/16. In round 15, selected residues in the HG3.14 variant were subjected to saturation mutagenesis (Supplementary Table 4a). Sites for mutagenesis were chosen either because they affected catalytic activity (Ser 89, His 90, Met 172, Thr 125 and Thr 265) or because they were near the substrate entry tunnel (Ser 89, His 90, Glu 46, Asn 47, Met 237, Phe 267, Trp 87, His 88, Trp 275 and Arg 276). Optimized codon sets were obtained by mixing two primers with different degenerate codons to minimize library size but maximize accessible amino acids³⁷. Because hits at positions 88 and 89/90 were too close to be efficiently combined by DNA shuffling, another focused library was prepared using very limited codon sets based on hits from round 15. Position 88 was randomized with an SMT codon (Asp, Pro, His and Ala), position 89 with MRT (Asn, His, Ser and Arg), and position 90 with a mix of YAC (Tyr and His) and TYT (Phe and Ser). This library was combined in a 1:4 molar ratio with a mix of HG3.7, 5 clones from round 14, and 14 clones from round 15 and shuffled according to standard protocols³³ to yield library 16.

Site-directed mutagenesis. The role of selected active site residues in HG3.17 was assessed by site-directed mutagenesis. For example, Asp 127, the intended catalytic base, was mutated to alanine and asparagine (primer sequences, Supplementary Table 3), whereas Gln 50, which emerged in round 4 of the directed evolution experiment, was mutated to alanine. Full-length genes were assembled using standard

library cloning primers (T7 terminator and T7 alternative, Supplementary Table 3). Ligation into the pET-11b(+) vector and subsequent transformation were performed as described above.

Screening rounds 1 to 10. Typically, 800–1,000 clones were screened per evolutionary round. HG3 gene libraries were transformed into BL21-Gold(DE3) cells and plated onto LB agar plates containing $150 \mu\text{g ml}^{-1}$ ampicillin. Single colonies were used to inoculate $180 \mu\text{l}$ LB medium containing $150 \mu\text{g ml}^{-1}$ ampicillin (LB-Amp¹⁵⁰) in a 96-well plate. After overnight incubation at 37°C , replica plates were generated. The overnight cultures were used to inoculate 2 ml LB-Amp¹⁵⁰ medium in 96-deep-well plates. The *E. coli* variants were grown for 5 h at 37°C before protein production was induced by addition of isopropyl- β -D-thiogalactoside (IPTG; 1 mM final concentration). The temperature was reduced to 18°C and protein production continued for 18 h. The cells were collected by centrifugation, the supernatant was discarded, and the cell pellets were frozen at -80°C for at least 1 h. After thawing, the cells were resuspended in $300 \mu\text{l}$ lysis buffer consisting of 50 mM sodium phosphate, pH 7.0, 100 mM NaCl and 1 mg ml^{-1} lysozyme. The cells were incubated for 1 h at room temperature before clearing the lysates by centrifugation. Catalytic efficiency was assayed by monitoring conversion of 5-nitrobenzisoxazole (250 μM final concentration) at 380 nm in a plate reader (ThermoFisher) at 27°C . The most active clones were confirmed by rescreening in triplicate. The clones with the largest increase in activity relative to the preceding round, typically corresponding to about 1% of the screened population, were picked from the replica plate for plasmid isolation, sequencing and further diversification.

Screening rounds 11 to 17. For rounds 11 to 17, HG3 libraries were cloned via NcoI and XhoI sites into expression vector pMG209, derived from pET-22b (Novagen)³⁸. At the 5' end of the insert, pMG209 encodes the periplasmic export sequence pelB in frame with the protein. For screening, BL21-Gold(DE3) cells were transformed with HG3 libraries. Single colonies were picked and used to inoculate round bottom 96-well plates containing $120 \mu\text{l}$ LB-Amp¹⁵⁰. After 18–22 h of growth at 30°C , protein production was induced with $40 \mu\text{l}$ of 1 mM IPTG in LB-Amp¹⁵⁰. Protein was expressed for 16–20 h at 18°C and cells were stored at 4°C until the catalytic assay was performed. Directly before the assay, cells were resuspended with a Lysidator 96 (Rainin). If necessary, cells were diluted in reaction buffer (50 mM sodium phosphate, pH 7, 100 mM NaCl) up to 100-fold. The assay reaction was performed with $20 \mu\text{l}$ of diluted or undiluted cell suspension in a total volume of $200 \mu\text{l}$ reaction buffer containing 125–250 μM 5-nitrobenzisoxazole and 1% MeCN in Nunclon coated flat-bottom plates (Nunc). Reaction progress was monitored in a plate reader (Molecular Devices) at 380 nm wavelength and 27°C . The most active clones were picked directly from the protein production plate for further characterization.

Protein production and purification. Selected variant(s) along the evolutionary trajectory were produced and purified for biochemical characterization. BL21-Gold (DE3) cells were transformed with the pET-11b(+) expression vector (Novagen) containing the gene of interest. To ensure monoclonality, single-colony streakouts were prepared to inoculate an overnight culture in LB-Amp¹⁵⁰. A 2-l Erlenmeyer flask containing 500 ml LB-Amp¹⁵⁰ was inoculated with 1 ml of the overnight culture and incubated until an attenuation ($D_{600 \text{ nm}}$) of 0.5–1 at 37°C . The temperature was then reduced to 18°C , and protein production induced with 250 μl IPTG. After 18–24 h, the cells were collected and cell pellets were frozen at -20°C . After thawing, pellets were resuspended in sonication buffer (50 mM Tris-HCl, pH 7.4, 500 mM NaCl) and $100 \mu\text{l}$ of a protease inhibitor cocktail (Sigma, P8849) added. Cells were lysed by addition of 0.5 mg ml^{-1} lysozyme and subsequent sonication. Cell debris was removed by centrifugation, and the soluble fraction was applied to a Ni-NTA resin (Qiagen) pre-equilibrated with sonication buffer. The samples were washed with the same buffer containing 20 mM imidazole before elution with 300 mM imidazole. The protein samples were washed into fast protein liquid chromatography (FPLC) buffer (20 mM sodium phosphate, pH 6, 20 mM NaCl) in an Amicon Ultra-15 centrifugal filter (10 kDa cut-off; Millipore) and then further purified by cation exchange chromatography (MonoS column, GE Healthcare) in the same buffer, eluting with a salt gradient (20–1,000 mM NaCl). The protein of interest typically eluted at a conductivity of 24 mS cm^{-1} . If necessary, proteins were concentrated using an Amicon Ultra-4 unit (10 kDa cut-off; Millipore). Protein concentrations were determined by measuring the absorbance at 280 nm using a calculated extinction coefficient (<http://expasy.org/tools/protparam.html>). Protein purity was confirmed by SDS-PAGE. Pure protein was stored in FPLC buffer at 4°C . The most active variant, HG3.17, retained full activity for several months under these conditions.

Circular dichroism spectroscopy. The far-ultraviolet spectrum of protein samples (5 μM) was measured in 50 mM sodium phosphate buffer, pH 7.0, and 100 mM NaCl at 20°C using an Aviv 202 spectropolarimeter (Aviv Associates). Thermal denaturation of the protein (5 μM) was monitored at 222 nm.

Mass spectrometry. For mass determination, protein samples were desalted on Illustra Nap-5TM columns (GE Healthcare) and measured in 0.1% acetic acid by

electrospray ionization–mass spectrometry (ESI–MS) on a Daltonics maXis ESI–Q–TOF mass spectrometer (Bruker). Mass spectra were deconvoluted using MaxEnt1 software. All masses corresponded to the expected sequence lacking the N-terminal methionine.

Substrate synthesis. 5-Nitrobenzisoxazole, 5-bromobenzisoxazole and 5-nitro-7-methoxybenzisoxazole were synthesized from the respective salicylaldehyde by first forming the oxime and then cyclizing with Ph_3P and DDQ³⁹. 5-Chlorobenzisoxazole, 5-cyanobenzisoxazole and 6-nitrobenzisoxazole were synthesized as previously described⁹.

Ultraviolet/visible spectroscopic assay. Reactions were initiated by adding enzyme (10 nM–10 μM) to the benzisoxazole substrate (50 μM –2 mM final concentration) in 50 mM sodium phosphate buffer, 100 mM NaCl, containing 10% methanol, pH 7.0. The pH of the buffer was measured after addition of methanol using a SenTix 81 pH electrode (Gerber Instruments). Product formation was monitored at the appropriate wavelength⁴⁰ in a Lambda 35 ultraviolet/visible (UV/Vis) spectrometer (PerkinElmer) at 27 °C. The slope before addition of protein was subtracted as background. Initial rates divided by catalyst concentration were plotted against substrate concentration, and k_{cat} and K_{m} values were determined by fitting the data to the Michaelis–Menten equation $v_0/[\text{catalyst}] = k_{\text{cat}}[S]/(K_{\text{m}} + [S])$. The uncertainty of the catalytic parameters k_{cat} and K_{m} was estimated by calculating the standard deviation of several independent measurements performed with different batches of protein. For the Kemp elimination products, extinction coefficients of 13,800 $\text{M}^{-1} \text{s}^{-1}$ at 408 nm for 5-nitro-7-methoxybenzisoxazole, and 4,930 $\text{M}^{-1} \text{s}^{-1}$ at 338 nm for 5-bromobenzisoxazole, were determined in alkaline aqueous solution. Leaving group $\text{p}K_{\text{a}}$ values were determined spectrophotometrically. For reactions at high substrate concentration (for example, Fig. 2d), product formation was monitored at 455 nm. After completion of the reaction, the sample was diluted 100-fold and product concentration determined from the absorbance at 380 nm as previously described⁴⁰.

Substrate solubility. K_{m} values of all HG3 variants were high compared to the solubility limit of the substrate. Solubility of 5-nitrobenzisoxazole under the assay conditions was quantified by diluting a 50 mM stock solution in methanol tenfold with assay buffer. The insoluble material was removed by centrifugation and the clear supernatant diluted into 100 mM NaOH solution. After complete conversion to 2-hydroxy-5-nitrobenzonitrile, the product was quantified spectrometrically. The solubility limit of 5-nitrobenzisoxazole under these conditions was determined to be 2.2 mM. For every kinetic assay, the background rate was plotted against the substrate concentration and a linear correlation served as evidence for the absence of non-ideal effects near the solubility limit.

K_{i} determination. HG3 variants (25 nM–1 μM final concentration, depending on their respective activity) were pre-incubated with varying concentrations of 6-nitrobenzotriazole (3) (0.001–1,000 μM final concentrations) in 50 mM Bis-Tris propane buffer, pH 6, containing 100 mM NaCl, 10% methanol and 1% dimethylsulphoxide (DMSO), and reactions were initiated by addition of 5-nitrobenzisoxazole (100 μM final concentration). Product formation was monitored as described previously. Half-maximum inhibitory concentration (IC_{50}) values were determined by curve fitting (Hill-slope model with the rate at infinite inhibitor concentration set to 0) and assumed to be equal to the K_{i} value at $[S] \ll K_{\text{m}}$ according to the Cheng–Prusoff equation⁴¹: $K_{\text{i}} = [\text{IC}_{50}]/(1 + [S]/K_{\text{m}})$.

pH-rate profile. The pH dependence of $k_{\text{cat}}/K_{\text{m}}$ was determined by measuring initial rates at 100 μM 5-nitrobenzisoxazole in different buffers. Acetate buffer was used from pH 4–5.5 and Bis-Tris propane buffer from pH 6–9.5, both containing 100 mM NaCl and 10% methanol. Enzyme was preincubated in the assay buffer for at least 5 min at 27 °C before initiation of the reaction by substrate addition. The rate of the uncatalysed reaction was measured with an identical sample without enzyme and subtracted. The apparent extinction coefficient for 5-nitro-2-hydroxybenzonitrile at 380 nm was corrected for the ionization state of the product using the formula $\epsilon_{380 \text{ nm}} = \epsilon_{\text{max}}/(1 + 10^{\text{p}K_{\text{a}} - \text{pH}})$; the product has a $\text{p}K_{\text{a}}$ of 3.98 and an ϵ_{max} of 15,800 $\text{M}^{-1} \text{cm}^{-1}$ when fully deprotonated.

Crystallization of HG3.17 complexed with a transition state analogue. Because crystallization trials with HG3.17 only yielded fine needles, a variant of the protein was chosen for detailed structural study in which two surface mutations (Asn47Glu and Asp300Asn) potentially involved in crystal contacts were reverted. This variant (100 μM) was incubated with 1 mM 6-nitrobenzotriazole (3) in 5 mM monobasic sodium phosphate, 100 mM sodium chloride, pH 7.0, containing 1% DMSO. The complex was crystallized by vapour diffusion in sitting drops comprising 200 nl of the enzyme–3 solution, 200 nl mother liquor (1.1 M ammonium sulphate, 100 mM sodium acetate, pH 5.9), and 50 nl diluted seed stock. HG3.7, a variant from the seventh round of evolution that readily crystallized, was used for cross-seeding. The seeds were generated with Seed Beads (Hampton Research) using pooled single crystals of HG3.7 grown in 1–2.5 M ammonium sulphate, 100 mM sodium acetate, pH 5–6.5. A 2.0 M ammonium sulphate solution was used as a crystal-stabilizing agent to collect the crystals and for dilution of seed stocks. Initial seed

stocks were diluted 1:1,000 to ensure that the crystals of the HG3.17 variant were not significantly contaminated with HG3.7. Under optimized sitting drop vapour diffusion conditions, one or two crystals were obtained per drop. Crystals were cryoprotected in mother liquor complemented with 20% (v/v) glycerol and flash frozen in liquid nitrogen.

Crystallographic methods. Crystallographic data were collected at 100 K using $\lambda = 1.00 \text{ \AA}$ at the Swiss Light Source (SLS) PXI-X06SA beamline (Paul Scherrer Institute, Villigen, Switzerland). The diffraction data were indexed and integrated with XDS⁴². Crystals belong to space group $P2_12_12_1$ ($a = 76.08 \text{ \AA}$, $b = 77.95 \text{ \AA}$, $c = 98.28 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$). The structure was solved by molecular replacement using PHASER⁴³ starting with the deposited model of HG2 (PDB accession code 3NYD)³. HG2, which was the direct precursor of HG3, has a serine residue at position 265 rather than a threonine as in HG3. The structure was refined in PHENIX⁴⁴ with manual corrections made in COOT⁴⁵. The asymmetric unit contains two independent chains of HG3.17 denoted 'A' and 'B'. The maximum-likelihood target function was used with optimized stereochemical and atomic-displacement parameter restraints for most of the refinement. In the final stage, the ligand as well as the catalytic residues Asp 127 and Gln 50 for the A-chain were refined unrestrained, using the least squares target function. The electron density reveals single conformations for most residues of the A-chain and two alternative conformations for the entire B-chain. The alternative B-chain conformations are modelled as a $1.2 \pm 0.6 \text{ \AA}$ rigid-body shift. To rule out the possibility that this observation was caused by improper space group assignment, the structure was refined in space group $P1$, all crystallographic subgroups of $P2_12_12_1$, as well as in space group $P2_1$. Because the alternative conformations were seen in all cases, we concluded that both B-chain conformations are not systematically distributed throughout the asymmetric units of lower symmetry space groups, as previously observed³ for the B-chain of HG2. Although crystals diffracted to higher resolution, data were cut off at 1.09 \AA because completeness dropped significantly below 50% in higher resolution shells. The backbone dihedral angles were distributed within the most favoured (98.7%) and additionally allowed (1.3%) regions of the Ramachandran map. Details of the refinement statistics are summarized in Supplementary Table 2.

Structural analysis. The length and width of the substrate entry tunnel were measured with the PyMOL plugin Caver 3.0 (ref. 46). The hydrogen-bonding interaction between Asp 127 and the transition state analogue was analysed by comparison with the optimal angles calculated for hydrogen-bonding interactions between acetamide dimers⁴⁷.

- Lassila, J. K., Privett, H. K., Allen, B. D. & Mayo, S. L. Combinatorial methods for small-molecule placement in computational enzyme design. *Proc. Natl Acad. Sci. USA* **103**, 16710–16715 (2006).
- Leung, D. W., Chen, E. & Goeddel, D. V. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**, 11–15 (1989).
- Stemmer, W. P. C. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl Acad. Sci. USA* **91**, 10747–10751 (1994).
- Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **9**, 1797–1804 (2008).
- Reetz, M. T. & Carballera, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols* **2**, 891–903 (2007).
- Kibbe, W. A. OligoCalc: An online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007).
- Kille, S. et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
- Gamper, M., Hilvert, D. & Kast, P. Probing the role of the C-terminus of *Bacillus subtilis* chorisate mutase by a novel random protein-termination strategy. *Biochemistry* **39**, 14087–14094 (2000).
- Iranpoor, N., Firouzabadi, H. & Nowrouzi, N. A novel method for the highly efficient synthesis of 1,2-benzisoxazoles under neutral conditions using the $\text{Ph}_3\text{P}/\text{DDQ}$ system. *Tetrahed. Lett.* **47**, 8247–8250 (2006).
- Kikuchi, K., Hannak, R. B., Guo, M. J., Kirby, A. J. & Hilvert, D. Toward bifunctional antibody catalysis. *Bioorg. Med. Chem.* **14**, 6189–6196 (2006).
- Cheng, Y.-C. & Prusoff, W. H. Relationship between the inhibition constant (K_{i}) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Adams, P. D. et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Chovancova, E. et al. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLOS Comput. Biol.* **8**, e1002708 (2012).
- Morozov, A. V., Kortemme, T., Tsemekhman, K. & Baker, D. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl Acad. Sci. USA* **101**, 6946–6951 (2004).

Structural basis for ligase-specific conjugation of linear ubiquitin chains by HOIP

Benjamin Stieglitz^{1*}, Rohini R. Rana^{1*}, Marios G. Koliopoulos¹, Aylin C. Morris-Davies¹, Veronique Schaeffer², Evangelos Christodoulou¹, Steven Howell¹, Nicholas R. Brown¹, Ivan Dikic² & Katrin Rittinger¹

Linear ubiquitin chains are important regulators of cellular signalling pathways that control innate immunity and inflammation through nuclear factor (NF)- κ B activation and protection against tumour necrosis factor- α -induced apoptosis^{1–5}. They are synthesized by HOIP, which belongs to the RBR (RING-between-RING) family of E3 ligases and is the catalytic component of LUBAC (linear ubiquitin chain assembly complex), a multisubunit E3 ligase⁶. RBR family members act as RING/HECT hybrids, employing RING1 to recognize ubiquitin-loaded E2 while a conserved cysteine in RING2 subsequently forms a thioester intermediate with the transferred or ‘donor’ ubiquitin⁷. Here we report the crystal structure of the catalytic core of HOIP in its apo form and in complex with ubiquitin. The carboxy-terminal portion of HOIP adopts a novel fold that, together with a zinc-finger, forms a ubiquitin-binding platform that orients the acceptor ubiquitin and positions its α -amino group for nucleophilic attack on the E3~ubiquitin thioester. The C-terminal tail of a second ubiquitin molecule is located in close proximity to the catalytic cysteine, providing a unique snapshot of the ubiquitin transfer complex containing both donor and acceptor ubiquitin. These interactions are required for activation of the NF- κ B pathway *in vivo*, and they explain the determinants of linear ubiquitin chain specificity by LUBAC.

Protein modification with ubiquitin is a key mechanism for the regulation of numerous cellular functions⁸. The transfer of ubiquitin onto a substrate is catalysed by E3 ligases, which can be classified into RING, HECT and RBR families^{9–13}. Two LUBAC subunits contain RBR domains. However, HOIP (HOIL-1L interacting protein) constitutes the catalytic centre, and its RBR domain-containing C-terminal region (HOIP_{RBR-C}) is sufficient to synthesize linear ubiquitin chains, regardless of E2 (Fig. 1a)^{6,14,15}. Although the RING1 domain of RBRs is assumed to be the primary binding site for E2s, we show that a HOIP construct containing only the catalytic cysteine-carrying RING2 plus a C-terminal extension (HOIP_{CBR-C}) still forms linear ubiquitin chains, roughly sevenfold more slowly than HOIP_{RBR-C} (Fig. 1b and Extended Data Fig. 1)^{7,15}. HOIP_{CBR-C} therefore constitutes the minimal catalytic core. We have determined the crystal structure of this catalytic core in its apo form at 2.4 Å and in complex with ubiquitin at 1.6 Å resolution (Extended Data Table 1). HOIP_{CBR-C} consists of seven α -helices and four zinc-binding modules, and its overall topology seems distinct from other structures (Fig. 1c–e and Extended Data Fig. 2). The helical region forms an elongated structural unit that acts as a platform, the ‘helical base’, to support the zinc-binding modules. The RING2 region of RBRs has recently been shown to adopt the in-between-ring (IBR) domain fold in auto-inhibited Parkin and HHARI (human homologue of ariadne)^{16–21}. This fold is preserved in the structure of active HOIP_{CBR-C} (Fig. 1c–e and Extended Data Fig. 3), and on the basis of this structural and high sequence conservation we suggest that it be renamed CBR (for ‘catalytic IBR’). The second zinc-binding site of the CBR of HOIP_{CBR-C} has a zinc-finger (ZF1) inserted between the second and third Zn²⁺-coordinating residues. A fourth zinc-binding

site is located between helices α 3 and α 4 (ZF2) and anchors the β -hairpin that is positioned close to the CBR.

In the HOIP_{CBR-C}-ubiquitin complex, ubiquitin makes contacts with residues from the helical base and ZF1 (Fig. 2 and Extended Data Fig. 4). No major conformational changes occur on complex formation, although some disordered regions of apo HOIP_{CBR-C} become ordered (Extended Data Fig. 2d). The ubiquitin bound by the helical base and ZF1 constitutes the acceptor ubiquitin: its α -amino group of M1 is located 3.5 Å away from the thioester-forming C885, poised for nucleophilic attack (Fig. 2). Strikingly, the C-terminal G76 of ubiquitin from a symmetry-related molecule is oriented such that its carboxylate points into the active site of HOIP_{CBR-C} sufficiently close to C885 to promote thioester formation. Thus, the molecular arrangement within the crystal lattice mimics the biologically relevant ubiquitin transfer complex with the donor and acceptor ubiquitin in an orientation consistent with linear chain synthesis (Fig. 2 and Extended Data Fig. 4). We found additional electron density in the active site, which we interpret as a Zn²⁺ ion. However, mass spectrometry, combined with structural and biochemical analysis, shows that ubiquitin chain synthesis is not a zinc-dependent process (Extended Data Figs 2b and 5).

HOIP_{CBR-C} employs residues from helices α 2 and α 6 in the helical base to contact T14, E16, D32 and K33 of the acceptor ubiquitin, whereas ZF1 rests against helix α 1 and the preceding loop (Fig. 3a and Extended Data Fig. 6a), ensuring that the N-terminal amino group is positioned closest to the active site and thus specifying linear chain synthesis. The side chain of M1 points away from the active site, indicating that its selection is driven stereochemically. To analyse the role of individual residues in ligase activity we used a combination of steady-state and single-turnover assays (Fig. 3c, d and Extended Data Fig. 6). Contributions from the helical base are crucial for ubiquitin chain synthesis, especially R935 and D936, but mutation of these residues does not impede thioester formation (Extended Data Fig. 6c). In ubiquitin the most severe effect was seen when K33 was mutated, and E16A and D32A showed decreased activity. No point mutation could be identified at the ZF1/ubiquitin interface that decreased chain synthesis; however, deletion of ZF1 resulted in an almost complete loss of activity (Fig. 3c). This impairment was not caused by protein misfolding, because ubiquitin-thioester formation was unaffected (Extended Data Fig. 6c). Instead, ZF1 deletion abolished transfer of the donor to the acceptor ubiquitin. These data suggest that the helical base constitutes the primary binding site for the acceptor ubiquitin, which is further supported by ZF1.

HOIP_{CBR-C} residues that contact the acceptor ubiquitin are located in a region termed the ‘linear ubiquitin chain determining region’ (LDD; residues 910–1082)¹⁵. Our structure now shows that this is not an independent ubiquitin-binding module, but together with the CBR it forms a superdomain that contacts donor and acceptor ubiquitin to create a platform that promotes linear chain synthesis.

Contacts between HOIP_{CBR-C} and the donor ubiquitin primarily involve the C-terminal tail of ubiquitin, which is guided towards the

¹Division of Molecular Structure, MRC National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK. ²Institute of Biochemistry II, Goethe University, School of Medicine, Theodor-Stern-Kai 7, D-60590 Frankfurt (Main), Germany.

*These authors contributed equally to this work.

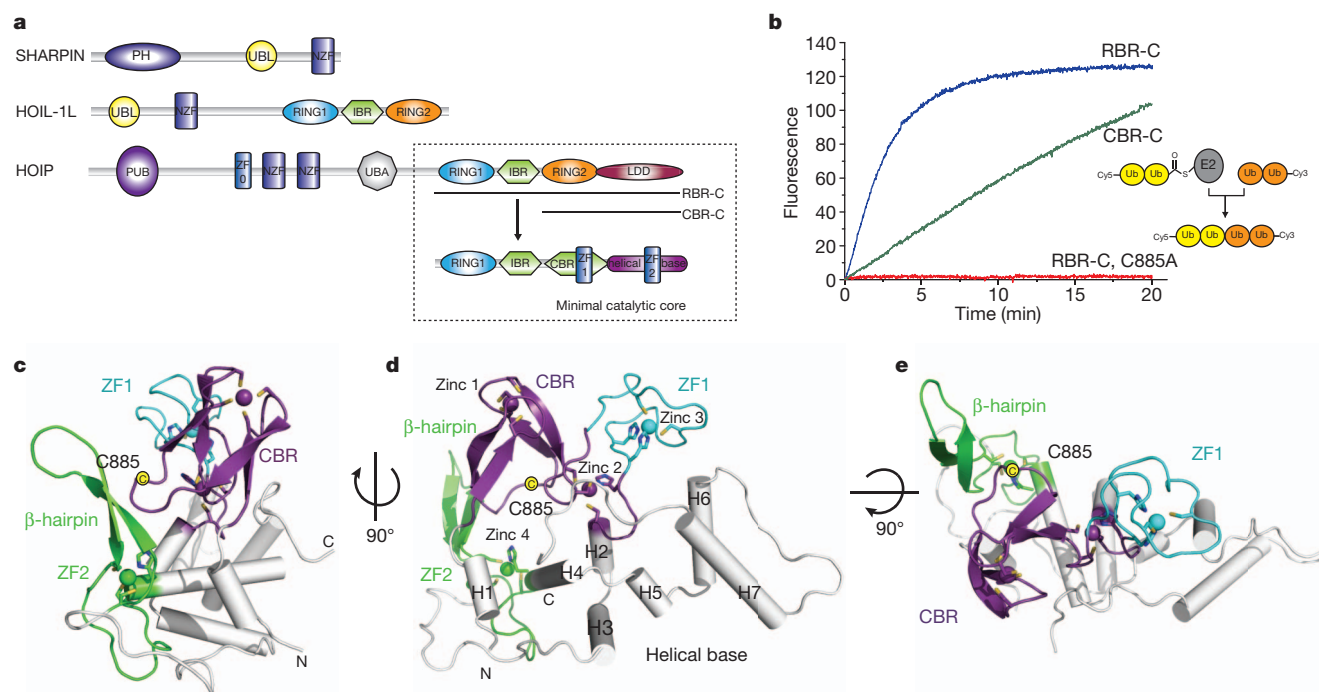


Figure 1 | Structure of the catalytic core of HOIP. **a**, Composition of LUBAC. SHARPIN, SHANK-associated RH domain-interacting protein; HOIL-1L, longer isoform of haem-oxidized iron-regulatory protein 2 ubiquitin ligase-1. Boxed: the crystallized catalytic core HOIP_{CBR-C}, biochemical assays employed HOIP_{RBR-C}. Below, diagram of new elements identified: CBR, ZF1, ZF2 and helical base. **b**, Single-turnover assays showing that lack of RING1

decreases activity 6.8-fold (HOIP_{CBR-C}, 0.050 min⁻¹, compared with HOIP_{RBR-C}, 0.341 min⁻¹). **c–e**, Ribbon representations of HOIP_{CBR-C} with the helical base in grey, CBR in purple, ZF1 in cyan, ZF2 and β -hairpin in green, Zn²⁺ ions as spheres, coordinating residues in ball-and-stick representation and the catalytic cysteine in yellow. The structure represents HOIP_{CBR-C} from the ubiquitin complex and includes regions disordered in the apo form.

active site through a channel created by the N-terminal antiparallel β -strands of the CBR and a β -hairpin formed by β F and β E (Figs 2 and 3b). These structural elements restrict tail mobility, ensuring that the carboxylate of G76 is located next to the catalytic cysteine. Contacts between Q974 and D983 in the β -hairpin and R72 and R74 from ubiquitin are crucial for donor ubiquitin binding (Fig. 3b–d and Extended

Data Fig. 6). The β -hairpin is largely disordered in the apo structure, suggesting that together with the CBR it could act as a flexible clamp locking the donor ubiquitin into place. Ubiquitin is further sandwiched by the N-terminal β -sheets of the CBR that form a hydrophobic pocket accommodating L73 and contacting L71. Mutation of either residue results in a severe loss of activity (Fig. 3c). Most of the hydrophobic

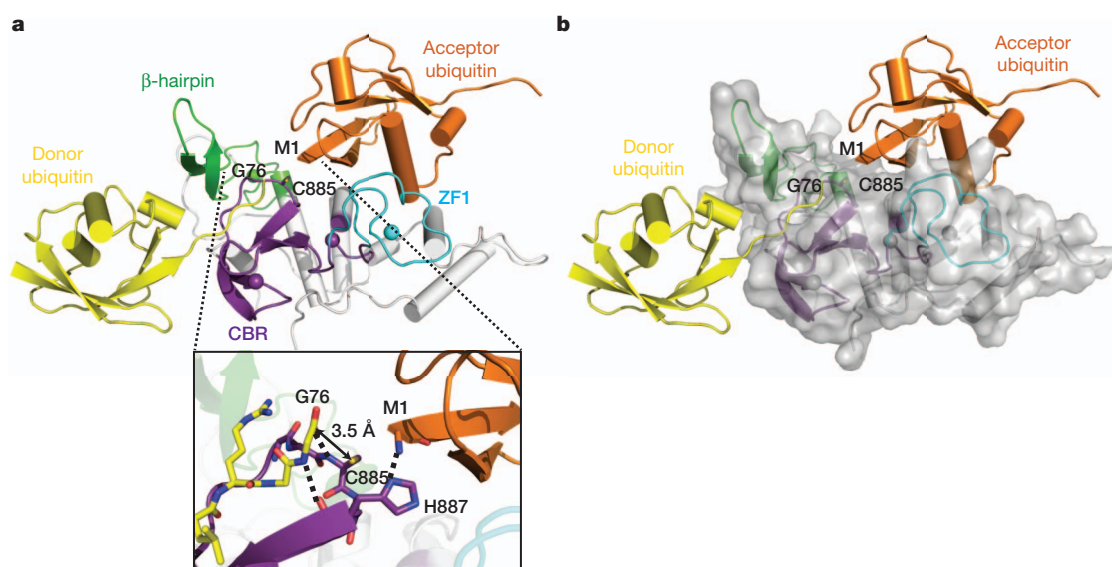


Figure 2 | The HOIP_{CBR-C}-ubiquitin transfer complex containing donor and acceptor ubiquitin. **a**, Ribbon representation of HOIP_{CBR-C} in complex with acceptor (orange) and donor (yellow) ubiquitin. HOIP_{CBR-C} is shown in the same orientation as in Fig. 1e. The positions of C885, donor G76 and acceptor M1 are indicated. Inset: contacts made by HOIP_{CBR-C} with donor and

acceptor ubiquitin. The arrow shows the proximity between G76 of the donor and S_Y of C885. **b**, The HOIP_{CBR-C}-ubiquitin complex with HOIP_{CBR-C} shown in a surface representation to emphasize the spatial relationship between the three molecules.

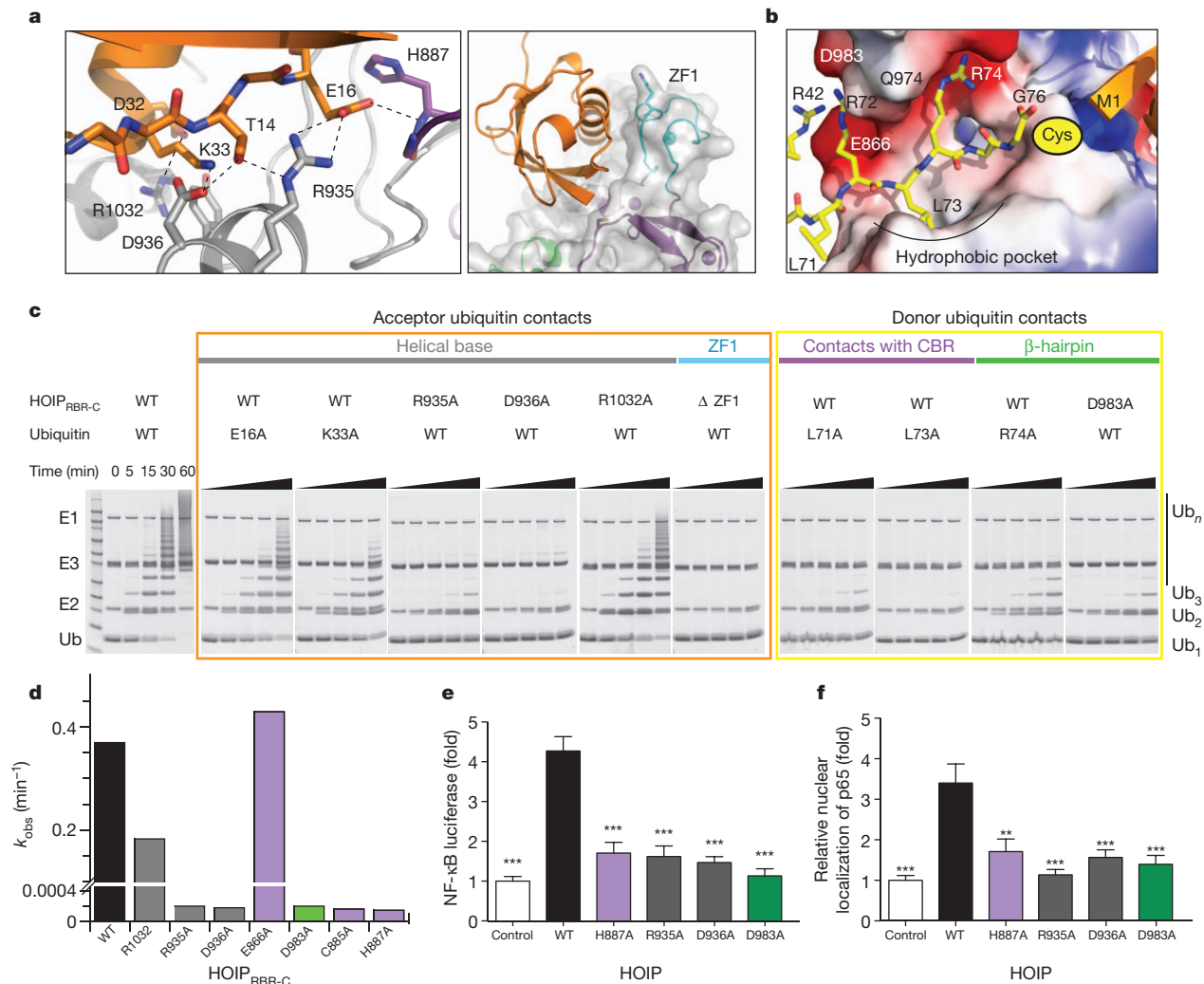


Figure 3 | Contacts between HOIP_{CBR-C} and ubiquitin required for ubiquitin transfer. **a**, Close-up of the HOIP_{CBR-C}/acceptor ubiquitin interface focusing on the helical base and ZF1. **b**, Details of the HOIP_{CBR-C}/donor ubiquitin (yellow) interface. Positions of C885 and acceptor ubiquitin M1 are indicated. **c**, Steady-state ubiquitination assays. Mutants that target the acceptor interface are boxed in orange, those with donor in yellow. Ub, ubiquitin; WT, wild type. **d**, Single-turnover assays to determine the rate of tetraubiquitin formation. **e**, Luciferase assays showing that the NF- κ B pathway

is not efficiently activated by HOIP mutants H887A, R935A, D936A and D983A in comparison with wild type. **f**, p65 translocations assay showing impaired p65 nuclear translocation on expression of HOIP ligase-deficient mutants. Three independent experiments were performed using triplicate samples. Results were analysed by ANOVA1 followed by Tukey post-tests. Error bars represent s.e.m. Two asterisks, $P < 0.01$; three asterisks, $P < 0.001$ compared with wild-type HOIP.

residues contacting L71 and L73 are conserved in CBRs, indicating that this mode of donor ubiquitin presentation to the substrate may be a general property of RBRs (Extended Data Fig. 3).

To confirm that the mechanism of linear ubiquitin chain synthesis identified in HOIP_{CBR-C} is maintained in LUBAC, we performed *in vitro* ubiquitination assays with the heterotrimeric complex. Although the activity of LUBAC was lower than that of isolated HOIP_{RBR-C}, possibly as a result of regulatory roles of ubiquitin-binding domains present in all three subunits, there was a strong correlation with the trends seen in ubiquitination assays using different ubiquitin mutants, indicating that domains outside HOIP_{RBR-C} do not affect chain linkage specificity (Extended Data Fig. 7). To validate our conclusions in a physiological context, we performed *in vivo* NF- κ B activation and p65 nuclear translocation assays employing HOIP mutants shown to affect interaction with the donor (D983A) and acceptor (R935A, D936A) ubiquitin. Consonant with our structural and biochemical data, these mutants decreased NF- κ B signalling and p65 nuclear translocation on overexpression (Fig. 3e, f) without impairing complex formation between LUBAC subunits (Extended Data Fig. 7).

Ubiquitin chain synthesis involves the nucleophilic attack of an amino group from a lysine or the N terminus of ubiquitin onto a ubiquitin-thioester formed by E2 or the HECT-type E3s. The reaction requires a general base to deprotonate the nucleophile and a mechanism to stabilize the transition state^{22–24}. In the HOIP_{CBR-C}-ubiquitin complex, H887 of HOIP_{CBR-C} forms a hydrogen bond with M1 of ubiquitin, indicating that it might be able to activate the incoming α -amino group or stabilize the transition state. To test its contribution to catalysis, we measured the activity of H887A, which was decreased more than 1,000-fold at 15 °C. This is due to lack of transfer to the acceptor ubiquitin rather than impairment of thioester formation (Figs 3d and 4a). Thus, H887 is not required for transthiolation from E2 to E3. To further characterize its function, we tested whether activity could be rescued at increased pH, which would aid deprotonation of the nucleophile. Indeed, ubiquitin transfer assays with H887A showed that activity could be restored at pH 9.0, although this required a higher temperature and higher substrate concentrations (Fig. 4b). Accordingly, enzyme activity was lost below pH 6.0, the approximate pK_a of an imidazole side chain (Extended Data Fig. 8). Taken together, these

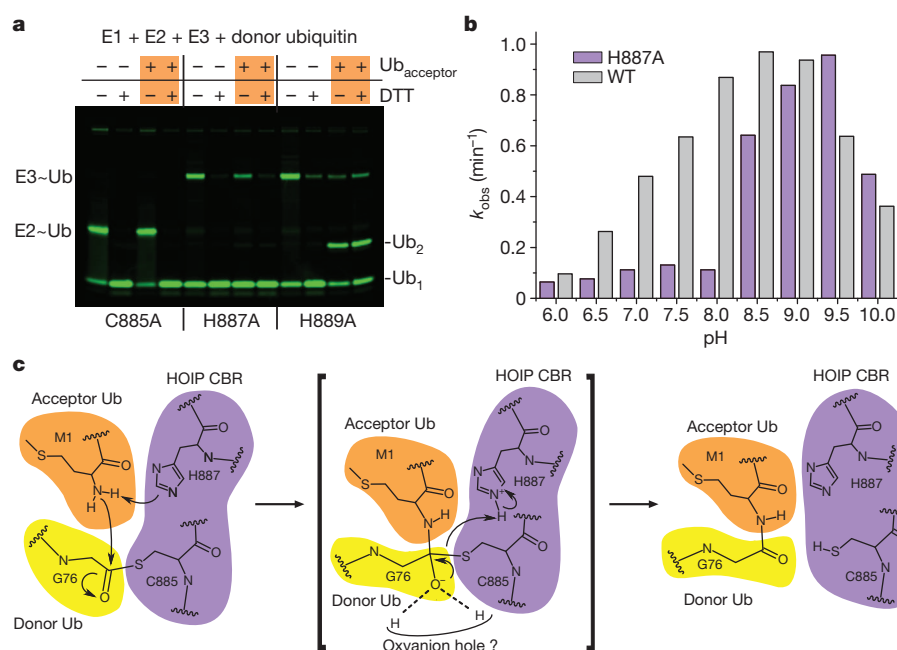


Figure 4 | HOIP H887 acts as the general base to activate the nucleophile. **a**, HOIP_{RBR-C} H887A and H889A (which shows wild-type activity) mutants can form a thioester (lanes 1 and 2 of each mutant), indicating that neither residue is involved in transthiolation from E2 to E3. However, H887A has lost the ability to transfer to a substrate to form diubiquitin. An orange rectangle indicates the presence of acceptor ubiquitin. DTT, dithiothreitol. **b**, Single-turnover assays of the pH dependence of tetraubiquitin formation, showing that the H887A mutant regains activity at higher pH at 25 °C. **c**, Proposed mechanism for ubiquitin transfer by HOIP.

observations support a model for ubiquitin transfer in which H887 acts as a general base to activate the nucleophile (Fig. 4c). This mechanism is maintained *in vivo* as indicated by the decrease in NF- κ B activation and p65 translocation by the H887A mutant (Fig. 3e, f). A histidine residue in this position is conserved in a number of RBRs and has recently been shown to be important for activity in Parkin and HHARI^{16,17,20}, suggesting that the mechanism of nucleophile activation may be conserved.

The structure of the HOIP_{CBR-C}-ubiquitin complex presented here provides the first insights into how an E3 ligase directs the synthesis of specific ubiquitin chains: a non-covalent ubiquitin-binding site orients the acceptor so that only the α -amino group of M1 is presented to the active site, in a similar manner to the mechanism used by linkage-specific E2s^{24–26} (Extended Data Fig. 9). M1 is part of a β -sheet and is less flexible than the ϵ -amino group of lysine, perhaps explaining why HOIP has evolved a single structural unit that integrates the CBR domain with the donor and acceptor ubiquitin-binding regions. Comparison of CBR structures from active HOIP with auto-inhibited Parkin and HHARI suggests that the overall mechanism of donor ubiquitin presentation is conserved in the RBR family (Extended Data Fig. 3c)^{16–21}. Further studies are now required to reveal the mechanism that promotes the formation of the active ligase complex and explain how chain linkage specificity is achieved in other RBR ligases.

METHODS SUMMARY

Proteins were expressed in *Escherichia coli* and purified by standard procedures. Steady-state ubiquitination, thioester formation and transfer assays were performed as described¹⁴. Diffraction data were collected at 100 K at Diamond Light Source, beamlines IO2 and IO4-1. The apo HOIP_{CBR-C} structure was solved by SAD, and the HOIP_{CBR-C}-ubiquitin complex was solved by molecular replacement.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 June; accepted 5 September 2013.

Published online 20 October 2013.

- Walczak, H., Iwai, K. & Dikic, I. Generation and physiological roles of linear ubiquitin chains. *BMC Biol.* **10**, 23 (2012).
- Tokunaga, F. *et al.* Involvement of linear polyubiquitylation of NEMO in NF- κ B activation. *Nature Cell Biol.* **11**, 123–132 (2009).
- Ikedo, F. *et al.* SHARPIN forms a linear ubiquitin ligase complex regulating NF- κ B activity and apoptosis. *Nature* **471**, 637–641 (2011).

- Tokunaga, F. *et al.* SHARPIN is a component of the NF- κ B-activating linear ubiquitin chain assembly complex. *Nature* **471**, 633–636 (2011).
- Gerlach, B. *et al.* Linear ubiquitination prevents inflammation and regulates immune signalling. *Nature* **471**, 591–596 (2011).
- Kirisako, T. *et al.* A ubiquitin ligase complex assembles linear polyubiquitin chains. *EMBO J.* **25**, 4877–4887 (2006).
- Wenzel, D. M., Lissounov, A., Brzovic, P. S. & Klevit, R. E. UBCH7 reactivity profile reveals Parkin and HHARI to be RING/HECT hybrids. *Nature* **474**, 105–108 (2011).
- Pickart, C. M. & Fushman, D. Polyubiquitin chains: polymeric protein signals. *Curr. Opin. Chem. Biol.* **8**, 610–616 (2004).
- Deshales, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78**, 399–434 (2009).
- Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. M. A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc. Natl Acad. Sci. USA* **92**, 2563–2567 (1995).
- Wang, M. & Pickart, C. M. Different HECT domain ubiquitin ligases employ distinct mechanisms of polyubiquitin chain synthesis. *EMBO J.* **24**, 4324–4333 (2005).
- Eisenhaber, B., Chumak, N., Eisenhaber, F. & Hauser, M. T. The ring between ring fingers (RBR) protein family. *Genome Biol.* **8**, 209 (2007).
- Wenzel, D. M. & Klevit, R. E. Following Ariadne's thread: a new perspective on RBR ubiquitin ligases. *BMC Biol.* **10**, 24 (2012).
- Stieglitz, B., Morris-Davies, A. C., Koliopoulos, M. G., Christodoulou, E. & Rittinger, K. LUBAC synthesizes linear ubiquitin chains via a thioester intermediate. *EMBO Rep.* **13**, 840–846 (2012).
- Smit, J. J. *et al.* The E3 ligase HOIP specifies linear ubiquitin chain assembly through its RING-IBR-RING domain and the unique LDD extension. *EMBO J.* **31**, 3833–3844 (2012).
- Trempe, J. F. *et al.* Structure of Parkin reveals mechanisms for ubiquitin ligase activation. *Science* **340**, 1451–1455 (2013).
- Wauer, T. & Komander, D. Structure of the human Parkin ligase domain in an autoinhibited state. *EMBO J.* **32**, 2099–2112 (2013).
- Spratt, D. E. *et al.* A molecular explanation for the recessive nature of Parkin-linked Parkinson's disease. *Nature Commun.* **4**, 1983, <http://dx.doi.org/10.1038/ncomms2983> (2013).
- Riley, B. E. *et al.* Structure and function of Parkin E3 ubiquitin ligase reveals aspects of RING and HECT ligases. *Nature Commun.* **4**, 1982, <http://dx.doi.org/10.1038/ncomms2982> (2013).
- Duda, D. M. *et al.* Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. *Structure* **21**, 1030–1041 (2013).
- Beasley, S. A., Hristova, V. A. & Shaw, G. S. Structure of the Parkin in-between-ring domain provides insights for E3-ligase dysfunction in autosomal recessive Parkinson's disease. *Proc. Natl Acad. Sci. USA* **104**, 3095–3100 (2007).
- Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Struct. Mol. Biol.* **13**, 491–499 (2006).
- Wu, P. Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22**, 5241–5250 (2003).

24. Wickliffe, K. E., Lorenz, S., Wemmer, D. E., Kuriyan, J. & Rape, M. The mechanism of linkage-specific ubiquitin chain elongation by a single-subunit E2. *Cell* **144**, 769–781 (2011).
25. Eddins, M. J., Carlile, C. M., Gomez, K. M., Pickart, C. M. & Wolberger, C. Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nature Struct. Mol. Biol.* **13**, 915–920 (2006).
26. Petroski, M. D. & Deshaies, R. J. Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* **123**, 1107–1120 (2005).

Acknowledgements We thank S. Strekopytov for inductively coupled plasma mass spectrometry experiments, I. Taylor for help with analytical ultracentrifugation, L. Haire for help with crystallization, S. Martin for circular dichroism spectroscopy, S. Caswell for help with cloning, C. Bayart for technical assistance, P. Walker for data collection, S. Smerdon for discussions and advice, and the Diamond Light Source for synchrotron access. This work was supported by the Medical Research Council (grant U117565398) and the Wellcome Foundation (grant 094112/Z/10/Z) to K.R., and the

European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) and ERC grant agreement 250241-LineUb to I.D.

Author Contributions B.S., R.R.R. and M.G.K. purified proteins and performed structural and biochemical analysis. A.C.M.-D. and N.R.B. purified proteins and conducted biochemical analysis. V.S. performed *in vivo* studies. E.C. produced expression plasmids. S.H. performed MS analysis. I.D. coordinated experimental work and contributed ideas. K.R. contributed to structural analysis and wrote the paper. All authors contributed to data analysis, experimental design and paper writing.

Author Information Coordinates and structure factors are deposited in the Protein Data Bank under accession codes 4LJQ (apo HOIP_{CBR-C} structure), 4LJO (wild-type HOIP_{CBR-C}-ubiquitin complex) and 4LJP (H889A HOIP_{CBR-C}-ubiquitin complex). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.R. (katrin.rittinger@nimr.mrc.ac.uk).

METHODS

Cloning, expression and protein purification. Cloning, expression and purification of Ube1, UbcH5A (UBE2D1), HOIP_{RBR-C} and mutants thereof, HOIP (residues 300–1072), HOIL-1L, SHARPIN and His₆-M1C-ubiquitin have been described¹⁴. HOIP_{CBR-C} was expressed and purified using the same procedure as for HOIP_{RBR-C}. Point mutations were generated using the QuikChange site-directed mutagenesis kit (Stratagene). The ZF1 of HOIP (residues 906–923) was deleted in HOIP_{RBR-C} and replaced with the sequence PG using Overlap Extension-PCR²⁷. Untagged ubiquitin and mutants were prepared according to ref. 28. Selenomethionine (SeMet)-substituted proteins were produced by standard procedures. Ubiquitin used for crystallization was purchased from Sigma and further purified by gel filtration. All plasmids were verified by DNA sequencing. Protein molecular mass was verified by electrospray ionization mass spectrometry. The fold of all proteins was analysed by circular dichroism spectroscopy.

Ubiquitination assays. Ubiquitination assays were performed using 1 μ M E1, 5 μ M UbcH5A, 5 μ M HOIP (or 5 μ M each: HOIP residues 300–1072, HOIL-1L and SHARPIN) and 200 μ M ubiquitin¹⁴. Reactions were incubated at 30 °C for 1 h and samples taken at 0, 5, 15, 30 and 60 min (HOIP_{RBR-C}) or 0.5, 1, 2 and 4 h (HOIP_{CBR-C}). Reactions were stopped by the addition of SDS sample buffer containing 40 mM *N*-ethylmaleimide. For LUBAC assays an additional precipitation step using 50 mM sodium acetate pH 4.0 at 60 °C was introduced. Samples were analysed by SDS-PAGE and visualized with Coomassie Brilliant blue.

Thioester formation and ubiquitin transfer assays. Labelling of His₆-Cys-ubiquitin with Cy5-Maleimide mono-Reactive Dye (GE Healthcare) and transfer assays were performed as described, with minor modifications¹⁴. Cy5-ubiquitin (1 μ M) was mixed with 2 μ M E1 and 1 mM ATP. After 5 min 10 μ M UbcH5A was added and after further 5 min 20 μ M HOIP_{RBR-C}. To monitor ubiquitin transfer, 10 μ M Ub-His₆ was added. Samples were taken before each addition and analysed by SDS-PAGE in the absence and presence of dithiothreitol.

Single-turnover fluorescence resonance energy transfer assays. UbcH5A was charged with Cy5-labelled linear di-ubiquitin and purified by gel filtration. E2~thioester (0.3 μ M or (for pH-dependent assays) 3.0 μ M) was mixed with 0.3 μ M or 3.0 μ M Cy3-labelled linear di-ubiquitin. After addition of 3.0 μ M of HOIP_{RBR-C} or HOIP_{CBR-C}, tetra-ubiquitin chain synthesis was observed by fluorescence resonance energy transfer between Cy3 and Cy5 using excitation and emission wavelengths of 540 and 670 nm, respectively. Samples were incubated at 15 °C in 50 mM HEPES pH 7.4, 150 mM NaCl, or at 25 °C for pH-dependent assays using phosphate (pH 6–8) and CHES (pH 8.5–10.0) buffers. Data were analysed by single exponential curve fitting. Data for constructs with very low activity were analysed by linear regression and a rate constant was calculated by dividing the obtained slope by the amplitude taken from the HOIP_{RBR-C} wild-type measurement. All measurements were performed in duplicate or triplicate; mean values are given.

Luciferase and p65 translocation assays. HeLa cells were transfected with pCMV-Flag-HOIP (wild-type, H887A, R935A, D936A and D983A), pcDNA5-haemagglutinin (HA)-SHARPIN, pNF- κ B-Luc (Stratagene) and β -GAL plasmids using Genejuice. After 36 h of transfection, lysates were prepared and subjected to luciferase assays in accordance with the manufacturer's protocol (Roche). Internal control was measured by β -galactosidase activity using its substrate (Roche). Three independent experiments were performed using triplicate samples in each experiment. For translocation assays HeLa cells were fixed and stained for p65 (sc-372; SantaCruz) and HA-SHARPIN (Covance, MMS-101P). The total number of cells showing nuclear staining of p65 was counted and normalized to the total number of cells expressing HA-SHARPIN to determine the percentage translocation. Three independent experiments were performed using triplicate samples in each experiment. Results were analysed by ANOVA1 followed by Tukey post-tests and are presented as means and s.e.m. Two asterisks, $P < 0.01$; three asterisks, $P < 0.001$ compared with wild-type HOIP.

Co-immunoprecipitation assays. HeLa cells were transfected with pCMV-Flag-HOIP, pcDNA5-HA-SHARPIN and pcDNA5-HA-HOIL-1L plasmids. After 36 h, Flag-HOIP immunoprecipitation was performed using anti-Flag M2 affinity gel (Sigma). The samples were subsequently probed for Flag-HOIP, HA-SHARPIN and HA-HOIL-1L.

Crystallization of apo HOIP_{CBR-C} and the ubiquitin complex. Crystallization trials with HOIP_{CBR-C} and its SeMet derivative were set up at 9.5 mg ml⁻¹ using an Oryx crystallization robot. Initial hits were optimized by sitting-drop vapour

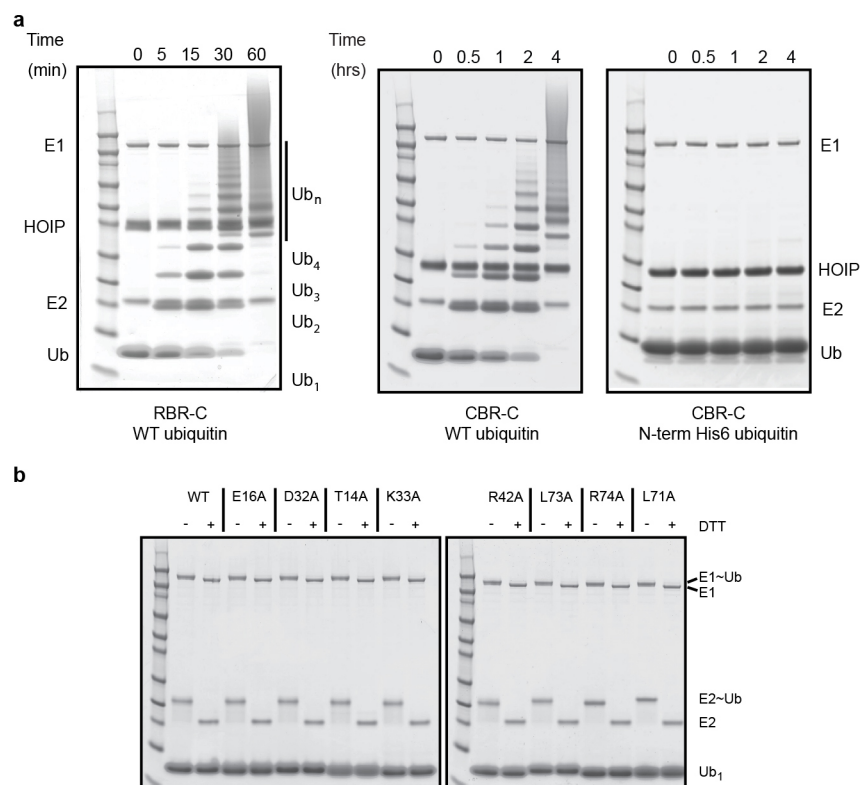
diffusion at 18 °C with a reservoir solution containing 100 mM Tris-HCl pH 8.5, 800 mM LiCl and 20% PEG 12000. Crystals were flash-frozen in the reservoir solution containing 30% glycerol. HOIP_{CBR-C} (0.6 mM) and ubiquitin were mixed at 1:2 and 1:3 molar ratios. Initial crystals were obtained in the Morpheus screen and optimized in hanging drops with reservoir solution containing 0.1 M carboxylic acids, buffer system 1 (0.1 M) pH 6.5, 20% PEG550 MME and 10% PEG20000. Crystals were flash-frozen in the reservoir solution. Crystals of mutant HOIP_{CBR-C} H899A with ubiquitin were grown under the same conditions as the wild-type protein.

Data collection and structure determination. Crystals of HOIP_{CBR-C} diffracted to 2.44 Å. A data set was collected on beamline IO2 ($\lambda = 0.9798$ Å) at the Diamond Light Source (Oxford, UK) and processed using XDS²⁹. The structure was solved by single-wavelength anomalous dispersion phasing using the SeMet derivative of HOIP_{CBR-C}. Heavy-atom search, density modification and initial model building were performed using Phenix AutoSol³⁰. Diffraction data for crystals of HOIP_{CBR-C} (wild type) and HOIP_{CBR-C} (H889A) in complex with ubiquitin were collected at beamlines IO2 ($\lambda = 1.282$ Å) and IO4-1 ($\lambda = 0.9163$ Å), respectively. Data were reduced using Xia2 from the CCP4 suite, and the structure of the complex was determined by molecular replacement in Phaser³¹ using the apo structure and ubiquitin (PDB 1UBQ) as search models. All models were iteratively improved by manual building in Coot³² and refined using REFMAC5 (ref. 33) and Phenix³⁰. The stereochemistry of the final models was analysed with Procheck. The model of apo HOIP_{CBR-C} has 94.8% of its residues in favoured regions, 4.2% in allowed regions and 1% outliers. The final models of wild-type HOIP_{CBR-C}-ubiquitin and HOIP_{CBR-C} H889A-ubiquitin have 97.6% and 95.5% of their residues in the favoured regions of the Ramachandran plot, respectively. Structural figures were prepared in Pymol.

Mass spectrometry. For zinc content analysis by native mass spectrometry, purified proteins were dialysed at 4 °C against 20 mM ammonium acetate pH 7.4. Molecular mass was determined by electrospray ionization (ESI) on a microTOFQ electrospray mass spectrometer (Bruker Daltonics, Coventry, UK). Protein was infused into the mass spectrometer at 3 μ l min⁻¹ using an electrospray voltage of 4.5 kV. Inductively coupled plasma mass spectrometry (ICP-MS) was used to determine the concentration of Ca and Zn (as ⁴⁴Ca and ⁶⁶Zn) in the protein samples using an Agilent 7700x instrument in helium (He) collision mode.

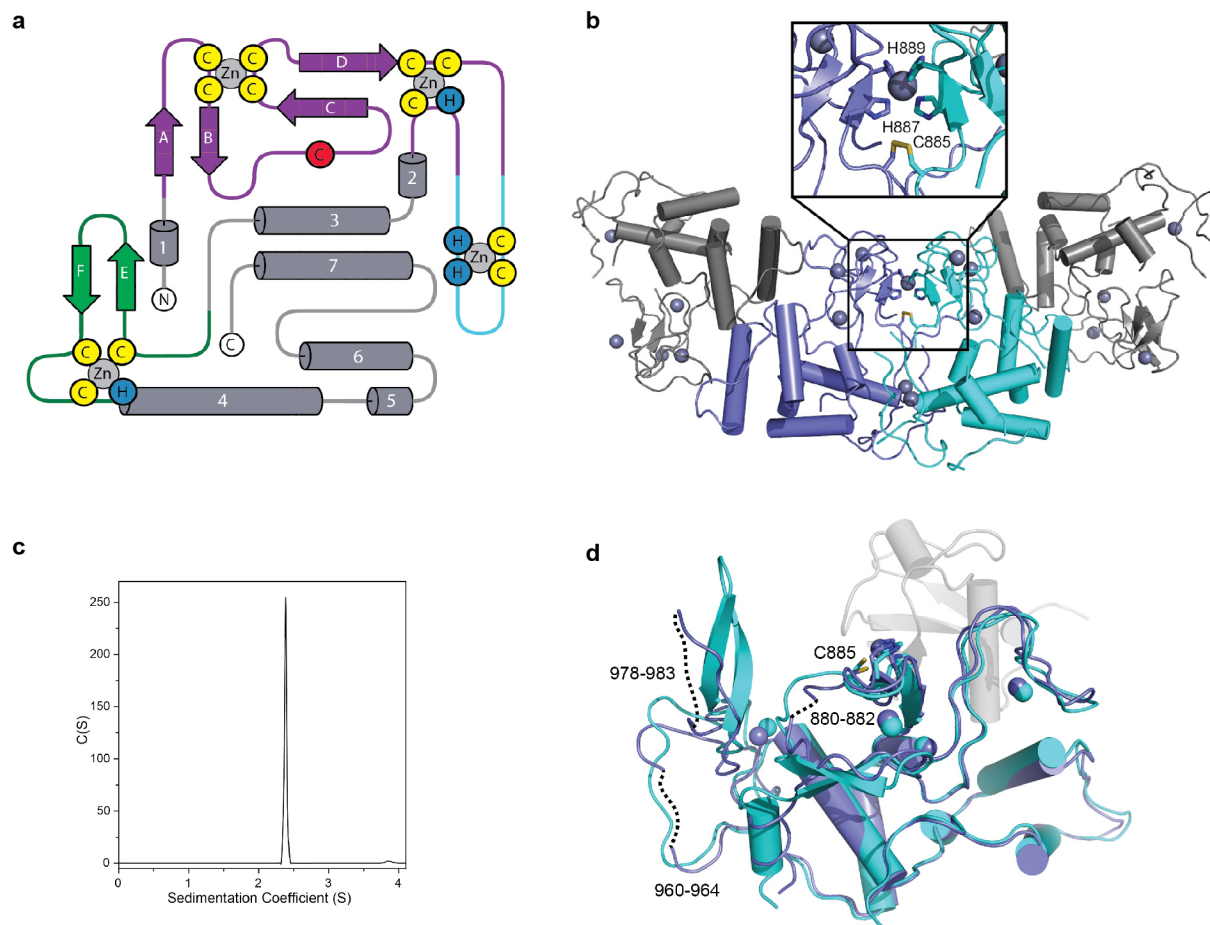
Analytical ultracentrifugation. Sedimentation velocity experiments were performed in a Beckman XL-I analytical ultracentrifuge. Samples were dialysed against the buffer blank, 20 mM Tris-HCl, 150 mM NaCl, 0.5 mM tris(2-carboxyethyl) phosphine pH 7.5. Centrifugation was performed at 50,000 r.p.m. (201,240g) and 293 K in an An50-Ti rotor at 125 μ M sample concentration. Data were analysed in terms of the size distribution function C(S) using the program SEDFIT³⁴.

27. Higuchi, R., Krummel, B. & Saiki, R. K. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351–7367 (1988).
28. Raasi, S. & Pickart, C. M. Ubiquitin chain synthesis. *Methods Mol. Biol.* **301**, 47–55 (2005).
29. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
30. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
31. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
32. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
33. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
34. Brown, P. H. & Schuck, P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.* **90**, 4651–4661 (2006).
35. Plechanovova, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H. & Hay, R. T. Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120 (2012).
36. Dou, H., Buetow, L., Sibbet, G. J., Cameron, K. & Huang, D. T. BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. *Nature Struct. Mol. Biol.* **19**, 876–883 (2012).
37. Prunedda, J. N. *et al.* Structure of an E3:E2~Ub complex reveals an allosteric mechanism shared among RING/U-box ligases. *Mol. Cell* **47**, 933–942 (2012).



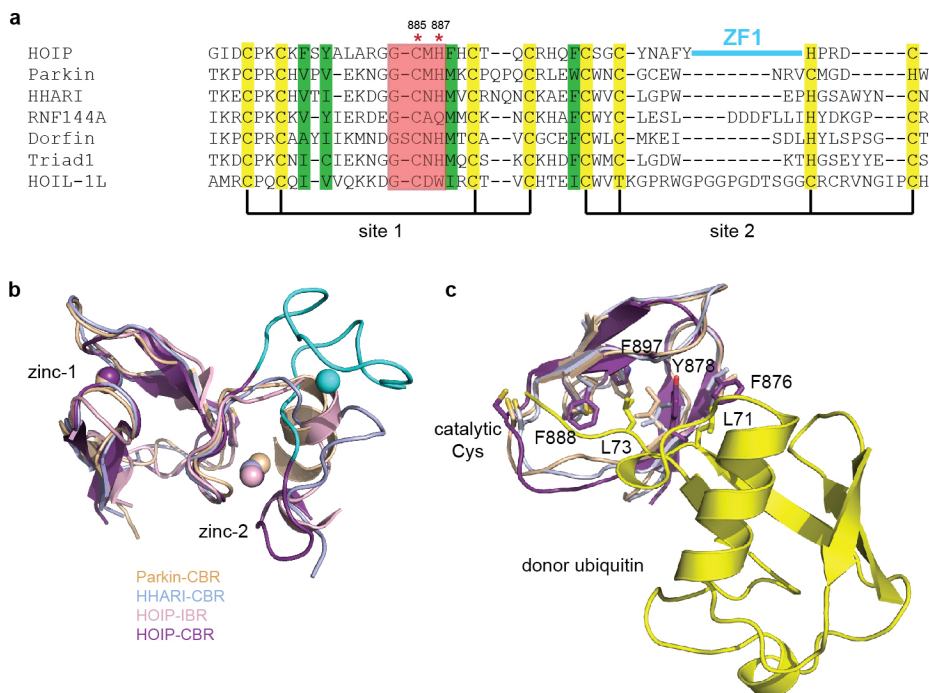
Extended Data Figure 1 | Comparison between the catalytic activities of the crystallized HOIP_{CBR-C} construct and the HOIP_{RBR-C} construct. **a**, Steady-state ubiquitination assays comparing the catalytic activity of the RBR-C construct of HOIP (HOIP_{RBR-C}) with the CBR-C construct (HOIP_{CBR-C}), showing that activity is reduced in HOIP_{CBR-C} with only short ubiquitin chains formed after 1 h. For this reason, all steady-state assays were performed with HOIP_{RBR-C}. To confirm that HOIP_{CBR-C} had retained the ability to specifically synthesize linear chains, a ubiquitination assay was performed with ubiquitin

containing an N-terminal His₆ tag that was no longer able to produce linear chains (right-hand gel). All gels were stained with Coomassie blue and converted to black and white. **b**, To ensure that any effects seen in ubiquitination assays were not due to inefficient loading of ubiquitin mutants onto E1 or E2 enzymes, the mutants were tested for their ability to form a thioester quantitatively with E1 and UbcH5A within 5 min. Samples were analysed on SDS gels and loaded in sample buffer in the absence or presence of dithiothreitol to detect the thioester.



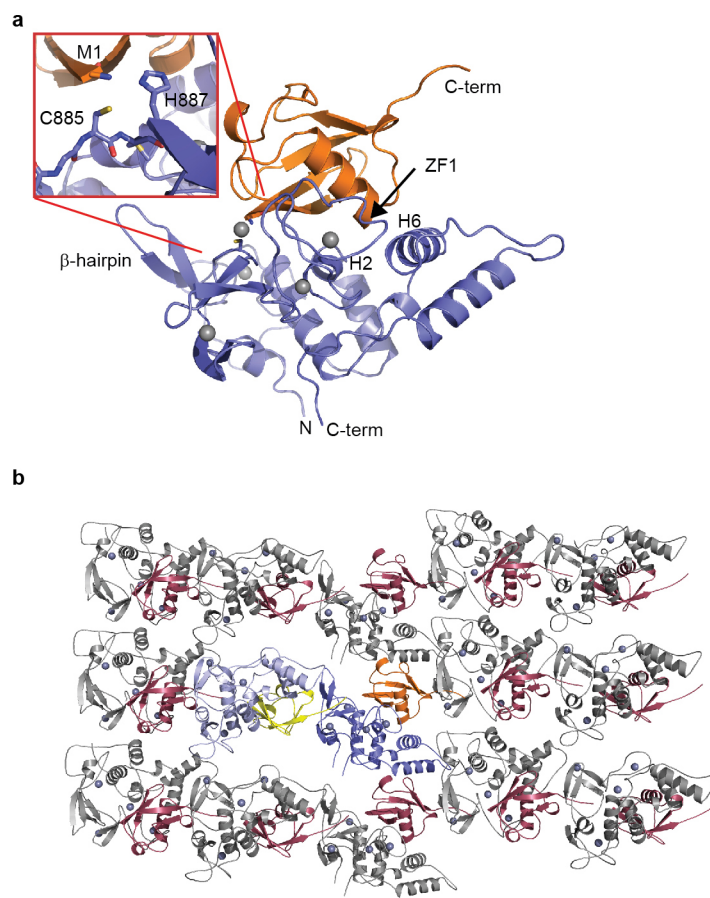
Extended Data Figure 2 | Topology of HOIP_{CBR-C}, content of the asymmetric unit of the apo crystals, solution behaviour of this fragment and comparison with the ubiquitin bound structure. **a**, Topology diagram of HOIP_{CBR-C} maintaining the same colour scheme as in Fig. 1. **b**, The apo HOIP_{CBR-C} structure contains four molecules of HOIP_{CBR-C} in the asymmetric unit that overlap with root mean squared deviation values of 0.8–1.0 Å. They are related by a local two-fold axis, and two molecules form a disulphide bond (shown in cyan and blue) and coordinate a fifth Zn²⁺ ion. Boxed: details of disulphide formation and zinc coordination by H887 and H889 of each monomer. Some of the loop regions in the apo structure are disordered. These

include: residues 866–868 in two monomers, 880–882 in all monomers, 960–964 in all monomers and 978–983, 973–982, 975–982 and 974–983 in the respective monomers. **c**, Sedimentation velocity run of the crystallized construct at a sample concentration of 125 μ M and derived C(S) distribution ($S_{app} = 2.38$; $S_{20,w} = 2.56$), indicating that it exists as a monomer in solution. **d**, Overlap of the apo (blue) and ubiquitin-bound (cyan) structures highlighting the regions that become ordered on complex formation (indicated by a dotted line) with ubiquitin (shown in grey). C885, H887 and H889 are shown as sticks and the Zn²⁺ ions as spheres.



Extended Data Figure 3 | Comparison of CBR domains from different RBR family members. **a**, Sequence alignment of different CBR domains. Conserved cysteine and histidine residues in the CBR involved in coordinating zinc 1 and zinc 2 are highlighted in yellow, and residues making hydrophobic contacts with the C-terminal tail of the donor ubiquitin are highlighted in green. The sequence forming ZF1 in HOIP has been removed for clarity and is indicated in cyan. The region around the catalytic C885 and H887, which are both crucial for catalytic activity, is highlighted in red. The glycine preceding C885 is conserved in other RBRs and might be important to allow the ubiquitin-loaded E2 access to the catalytic cysteine. The sequence variation of site 2 is much higher than for site 1, and in particular the sequence between zinc-coordinating residues 2 and 3, which in HOIP accommodates ZF1, varies significantly in length. **b**, This sequence variation is reflected in the structures of CBRs from

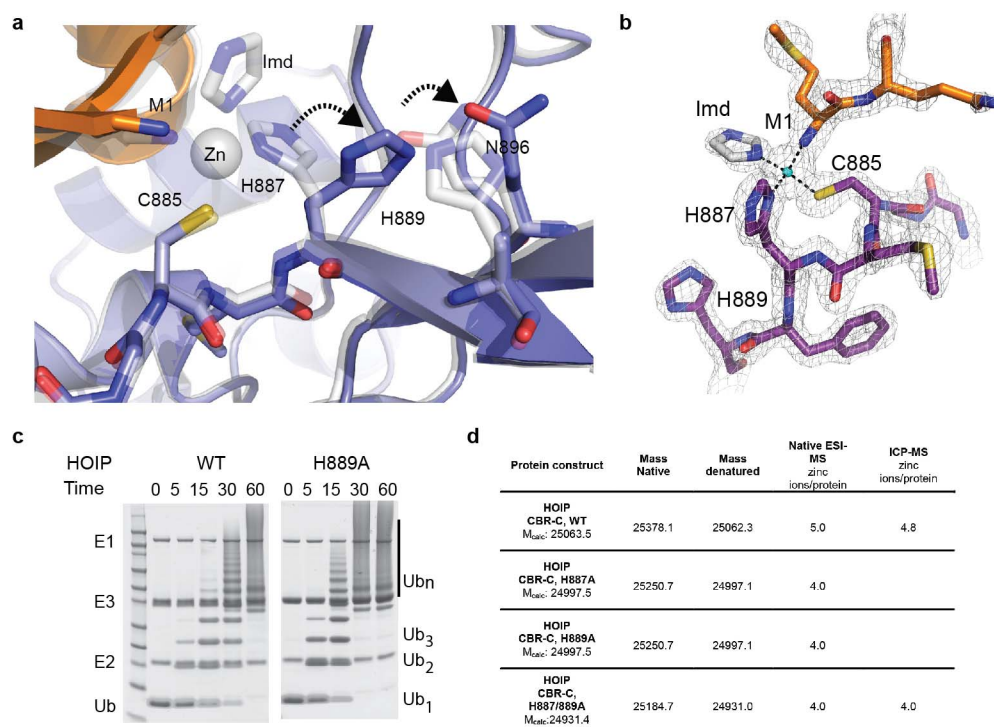
HOIP, Parkin (PDB 4K7D)¹⁶ and HHARI (PDB 4KC9)²⁰, which overlap well in site 1 and the following first two zinc-coordinating residues of site 2 but subsequently diverge significantly. Nevertheless, the positions of zinc 2 overlap well. The IBR of HOIP (pink; PDB 2CT7)²¹ is shown for comparison in the overlap. The ZF1 is shown in cyan and the Zn^{2+} ions are shown as spheres. **c.** HOIP_{CBR-C} in complex with the donor ubiquitin, overlapped with the CBRs of Parkin and HHARI, showing conserved hydrophobic residues that contact L71 and L73 of ubiquitin in ball-and-stick representation as well as the catalytic cysteine of the ligase. The C-terminal portion of the CBR (site 2) has been omitted for clarity. This overlap shows clearly that the C-terminal tail of the donor ubiquitin could be accommodated in a similar manner in Parkin and HHARI.



Extended Data Figure 4 | The HOIP_{CBR-C}-ubiquitin complex in the asymmetric unit and arrangement in the crystal lattice. **a**, The asymmetric unit contains one molecule of HOIP_{CBR-C} (blue) and one molecule of ubiquitin (orange). The ubiquitin bound to HOIP_{CBR-C} represents the acceptor ubiquitin; its α -amino group of M1 is located in close proximity to the thioester-forming C885, which is shown in ball-and-stick representation and the Zn²⁺ ions are shown as grey spheres, including the Zn²⁺ ion found in the active site. Inset:

details of the active site highlighting the proximity of M1 and C885 and H887. The Zn²⁺ ion found in the active site has been removed for clarity.

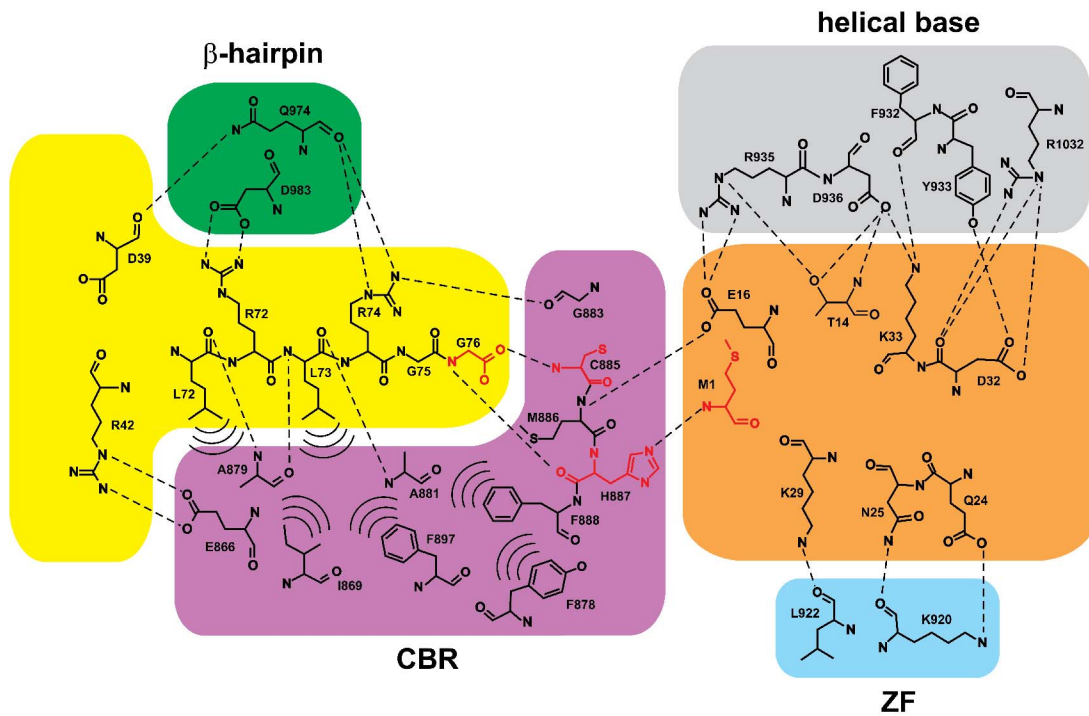
b, HOIP_{CBR-C} and ubiquitin that constitute the asymmetric unit are shown in blue and orange, respectively. A symmetry-related complex that contributes the donor ubiquitin is shown in yellow (ubiquitin) and light blue (HOIP). All other complexes in the lattice are shown in grey (HOIP) and red (ubiquitin).



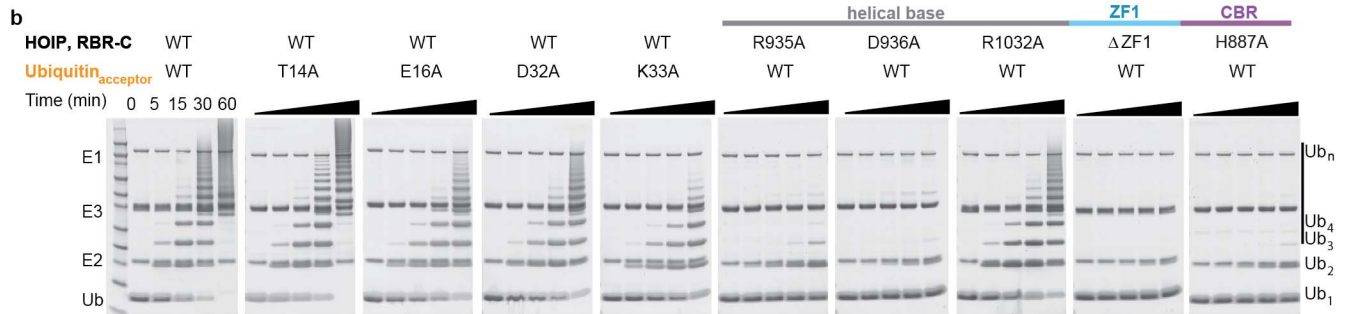
Extended Data Figure 5 | Active site arrangement including coordination of the fifth Zn^{2+} ion and ubiquitination assays with wild-type HOIP_{RBR-C} and the H889A mutant. **a**, We found residual electron density in the active site of the wild-type HOIP_{CBR-C}-ubiquitin complex (shown in transparent grey), which adopts a tetrahedral coordination and which we interpret as a Zn^{2+} ion. This Zn^{2+} is coordinated by the catalytic cysteine, H887, the α -amino group of ubiquitin and an imidazole from the crystallization solution (Imd). The observation that apo and substrate-bound HOIP_{CBR-C} contain a metal ion close to the active site prompted us to investigate a possible role in catalysis. Metal binding was examined by native electrospray mass spectrometry and ICP-MS, which indicated the presence of roughly five Zn^{2+} ions in the wild-type protein. Mutation of either of the histidines H887A and H889A, which coordinate the additional zinc in the apo structure, decreased the number to four. The structure of the H889A mutant in complex with ubiquitin (in blue and orange, respectively) lacks additional electron density in the active site, while retaining full catalytic activity, indicating that the catalytic step is not metal-dependent. Instead we believe that high reactivity of the active site induces disulphide bond

formation and subsequent zinc coordination across the interface during crystallization of apo HOIP_{CBR-C}, whereas in the substrate-bound complex the active site cysteine itself coordinates a metal ion with the help of an imidazole molecule from the crystallization buffer. **b**, Sigma-A weighted omit map contoured at 1.5σ showing the Zn^{2+} ion in the active site and its coordination by the α -amino group of ubiquitin M1, C885 and H887 from HOIP_{CBR-C}, and an imidazole molecule from the crystallization buffer. **c**, Ubiquitination assays comparing the activity of wild-type HOIP_{RBR-C} and the H889A mutant that no longer coordinates a fifth Zn^{2+} ion but retains full catalytic activity. **d**, The molecular mass of each construct listed (in daltons) was determined in its native and denatured forms by ESI-MS. The difference in mass of native and denatured proteins was used to calculate the number of Zn^{2+} ions present (63.4 Da per Zn^{2+}). The calculated mass of the constructs under investigation contains the additional sequence GPG that remains after removal of the glutathione S-transferase tag with PreScission protease. Metal analysis by ICP-MS did not reveal the presence of significant amounts of metal ions apart from Zn^{2+} .

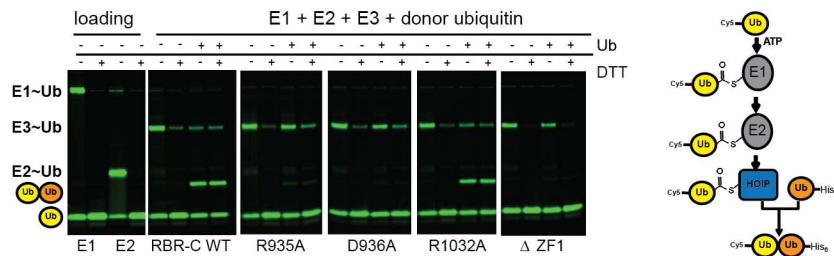
a



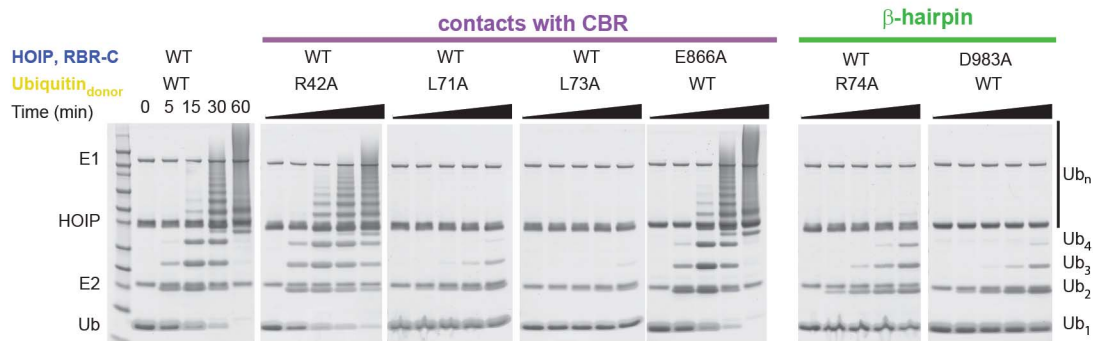
b



c



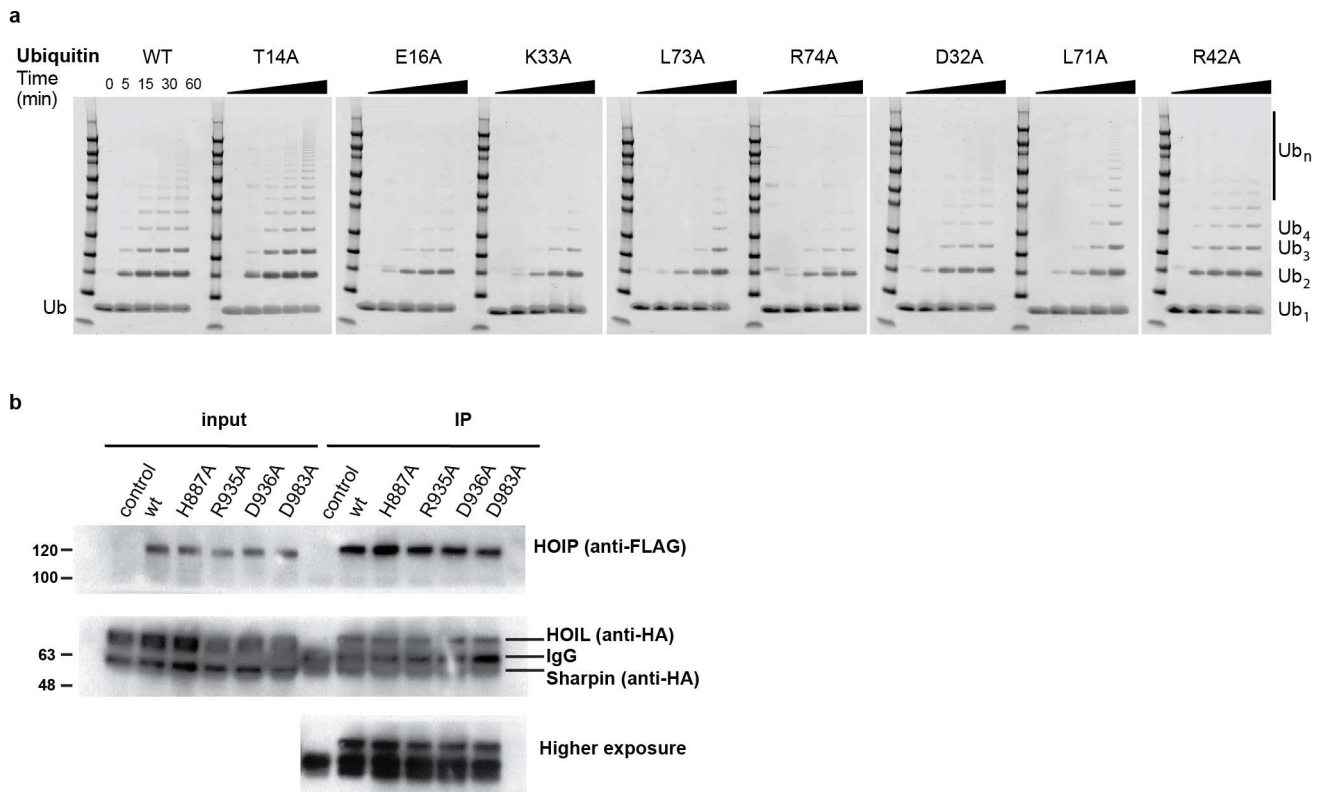
d



Extended Data Figure 6 | Diagram of the HOIP_{CBR-C}-ubiquitin complex interface and complete set of steady-state ubiquitination assays performed with HOIP_{RBR-C}.

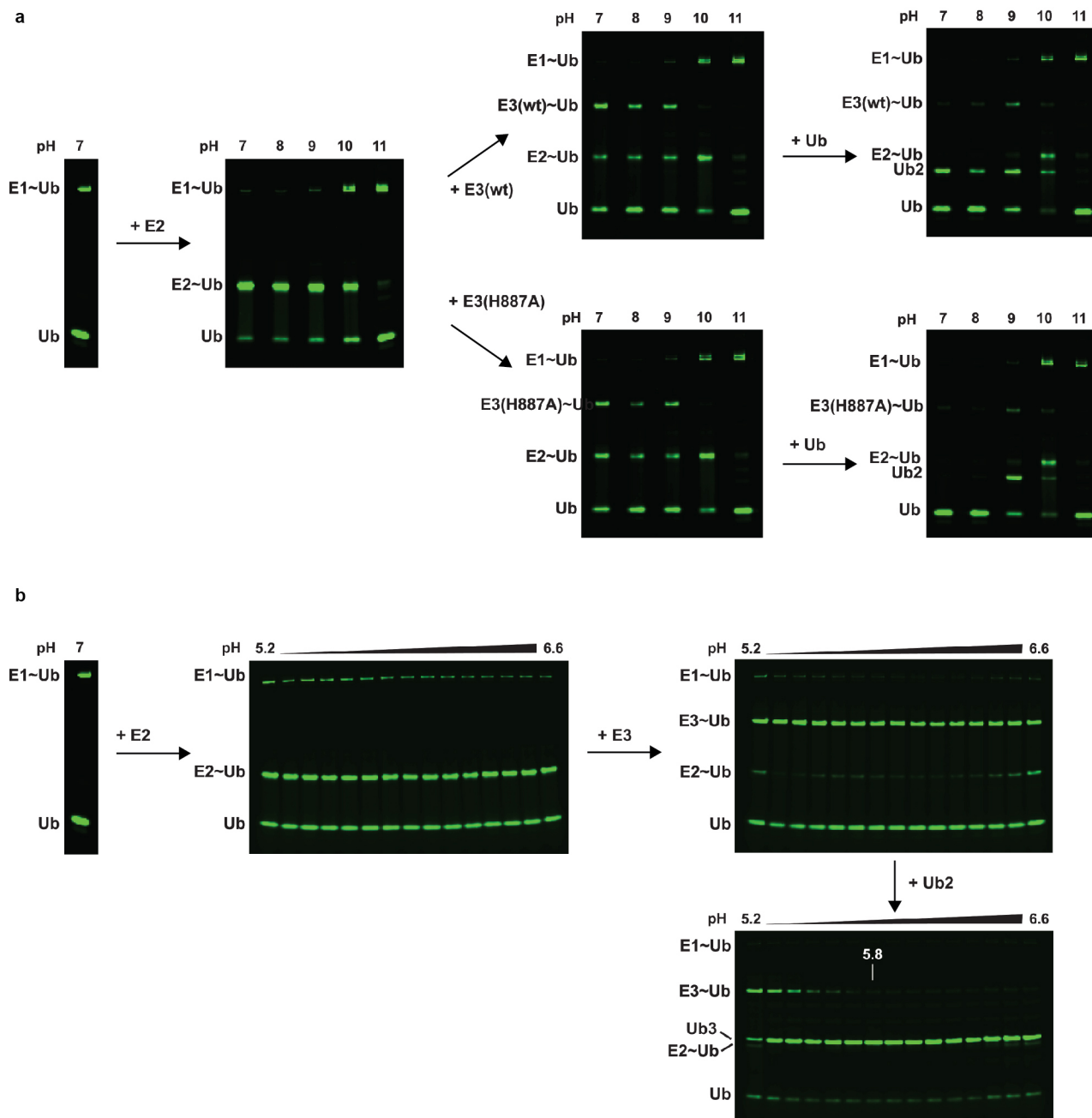
a, Diagram of the interface between HOIP_{CBR-C} and donor and acceptor ubiquitin, respectively. The colour scheme used in Fig. 1 has been maintained, with a cut-off of 3.5 Å for polar and 4.0 Å for hydrophobic interactions. The C terminus of the donor ubiquitin is oriented towards the catalytic cysteine through multiple interactions with HOIP. The carbonyl of G76 forms a hydrogen bond with the C885 backbone amide and is 3.5 Å distant from its S_γ. The backbone NH of G76 makes further contact with the loop carrying C885, whereas an extended conformation for the rest of the tail is maintained by interactions with the CBR and β-hairpin. This arrangement is reminiscent of RING ligases that lock the E2~Ub in a folded-back conformation^{35–37}, indicating that this might be a general mechanism to activate a ubiquitin thioester intermediate. **b**, The interface between HOIP and the acceptor ubiquitin includes residues from the helical base of HOIP that interact with T14, E16, D32 and K33. Mutation of T14 in HOIP_{RBR-C} to alanine has only

a modest effect on the steady-state synthesis of linear ubiquitin chains, whereas E16 and D32 show significantly decreased activity; the strongest effect was seen with the K33A mutant. In HOIP_{RBR-C}, mutation of R935 and D936 almost completely abrogates activity, whereas the R1032A mutant has only a minor effect. **c**, E3-thioester formation and ubiquitin transfer assays. The diagram illustrates the set-up: lanes 1–4, donor Cy5-ubiquitin loading onto E1 and E2, followed by addition of wild-type HOIP_{RBR-C} or mutants to form an E3~thioester (lanes 1 and 2 of each experiment). To monitor ubiquitin transfer, a C-terminally blocked ubiquitin was added (lanes 3 and 4 of each experiment). **d**, Mutations that disrupt the interaction of the C-terminal tail of the donor ubiquitin with the conserved region of the CBR and the β-hairpin strongly decrease the ability to produce linear ubiquitin chains. In contrast, mutation of E866, which is not conserved in other RBR family members, and its contact R42 has no effect on ubiquitin chain synthesis. Some of the gels included in this figure are also shown in Fig. 3c and are included for comparison.



Extended Data Figure 7 | *In vitro* ubiquitination assays with heterotrimeric LUBAC and *in vivo* co-immunoprecipitation of HOIP mutants with SHARPIN and HOIL-1L. a, *In vitro* ubiquitination assays with heterotrimeric LUBAC, showing that the overall activity of LUBAC is lower than that of isolated HOIP_{RBR-C}, possibly as a result of regulatory roles of the ubiquitin-binding domains that are present in all three LUBAC subunits. Nevertheless, the trends observed with isolated HOIP_{RBR-C} are conserved with LUBAC: those mutations that had only a minor or no effect on chain synthesis (T14A and

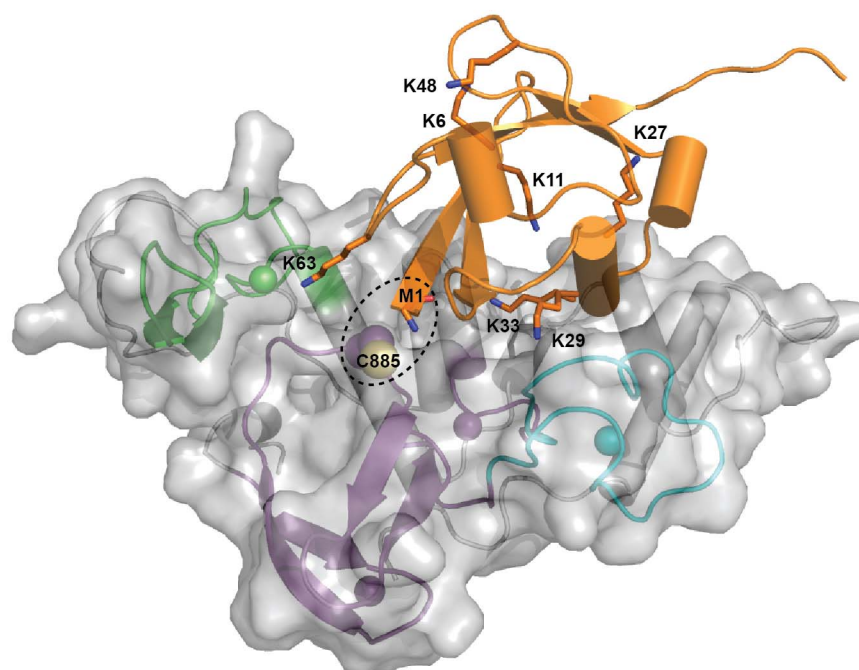
R42) show the same behaviour, whereas the E16A, K33A, L71A, L73A and R74A mutants show a significant decrease in ubiquitin chain synthesis. **b, Co-immunoprecipitation assays show that mutations in HOIP that interfere with the binding of donor or acceptor ubiquitin and interfere with ubiquitin chain synthesis have no effect on complex formation between LUBAC subunits, and hence any effects seen in NF- κ B activation and p65 translocation assays are not due to impaired complex formation.**



Extended Data Figure 8 | pH dependence of ubiquitin transfer with wild-type HOIP_{RBR-C} and the H887A mutant. Ube1 (0.5 μ M) was charged with 1 μ M His₆-Cy5-ubiquitin (Ub) using 5 mM ATP at pH 7. **a**, The pre-charged E1-ubiquitin thioester (E1~Ub) was subsequently mixed with 10 μ M UbcH5A and incubated for 5 min under different buffer conditions ranging from pH 7 to pH 11. Complete ubiquitin transthiolation from the E1 onto UbcH5A (E2~Ub) can be observed at pH 7–9, is impaired at pH 10 and abolished at pH 11. Wild-type HOIP_{RBR-C} (20 μ M; E3(wt), top row) or HOIP_{RBR-C} H887A (20 μ M; E3(H887A), bottom row) were added to each sample and incubated for a further 5 min. Under these conditions, a thioester intermediate for both wild-type (E3(wt)~Ub) and mutant HOIP_{RBR-C} (E3(H887A)~Ub) can be detected at pH 7–9. All samples were finally mixed with 10 μ M C-terminal His₆-tagged ubiquitin (Ub). HOIP_{RBR-C} wild-type catalyses the formation of di-ubiquitin (Ub₂) at pH 7–9 and to some extent at pH 10. In contrast, product formation is

absent at pH 7 and pH 8 for HOIP_{RBR-C} H887A, indicating that histidine 887 is required for catalysis under physiological pH conditions. The assay was performed in 5 mM MgCl₂, 150 mM NaCl and 200 mM buffer (HEPES pH 7.0, HEPES pH 8.0, CHES pH 9.0, CHES pH 10.0 and CAPS pH 11.0).

b, The pre-charged E1-ubiquitin thioester (E1~Ub) was mixed with 10 μ M UbcH5A (E2) and incubated for 5 min in 150 mM sodium acetate buffer ranging from pH 5.2 to pH 6.6 with 0.1 increments. All samples display the same amount of charged UbcH5A (E2~Ub). Similarly, the addition of 20 μ M HOIP_{RBR-C} (E3) shows formation of the thioester charged intermediate (E3~Ub) to the same extent. Each sample was mixed with 10 μ M C-terminal His-tagged diubiquitin (Ub₂) as acceptor to allow product formation (Ub₃). The discharge of E3-thioester ubiquitin onto the acceptor is clearly impaired at pH values below 5.8. Gels were run under non-reducing conditions.



Extended Data Figure 9 | The structural basis of chain linkage specificity. Surface representation of HOIP_{CBR-C} with the acceptor ubiquitin in orange ribbon representation, indicating the position of all seven lysine residues present in ubiquitin, as well as the N-terminal methionine and the catalytic

cysteine, C885, of HOIP. The figure clearly shows that the α -amino group of methionine 1 is closest to the active-site cysteine, explaining the specificity for linear chains.

Extended Data Table 1 | Data collection and refinement statistics

	Apo HOIP	HOIP WT/ubiquitin	HOIP H889A/ubiquitin
Data collection			
Space group	P1	P 3 ₁	P 3 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	44.21, 47.77, 111.14	45.95, 45.95, 133.01	46.0, 46.0, 133.37
α , β , γ (°)	101, 90, 99	90, 90, 120	90, 90, 120
Resolution (Å)	29.61 (2.44-2.59)*	44.34-1.56 (1.60-1.56)*	38.17-2.15 (2.21-2.15)*
<i>R</i> _{sym} or <i>R</i> _{merge}	0.129 (0.576) *	0.039 (0.581) *	0.094 (0.614) *
<i>I</i> / σ <i>I</i>	10.07 (2.64) *	12.2 (2.1) *	9.3 (2.4) *
Completeness (%)	91.7 (71.0) *	99.2 (98.8) *	99.5 (99.5) *
Redundancy	3.33 (2.80) *	3.4 (3.4) *	5.1 (5.4) *
Refinement			
Resolution (Å)	29.61-2.44	44.34-1.56	38.17-2.15
No. reflections	30946	84025	33376
<i>R</i> _{work} / <i>R</i> _{free}	20.64/24.27	18.12/21.24	17.57/21.61
No. atoms			
Protein	5849	2306	2292
Zn ²⁺	18	6 (+imidazole)	4
Water	46	222	120
B-factors			
Protein	44.61	38.7	50.3
Zn ²⁺	43.88	30.6	37.0
Water	36.88	41.9	42.7
R.m.s deviations			
Bond lengths (Å)	0.0130	0.006	0.008
Bond angles (°)	1.790	1.032	1.115

One crystal was used for each of the data sets.

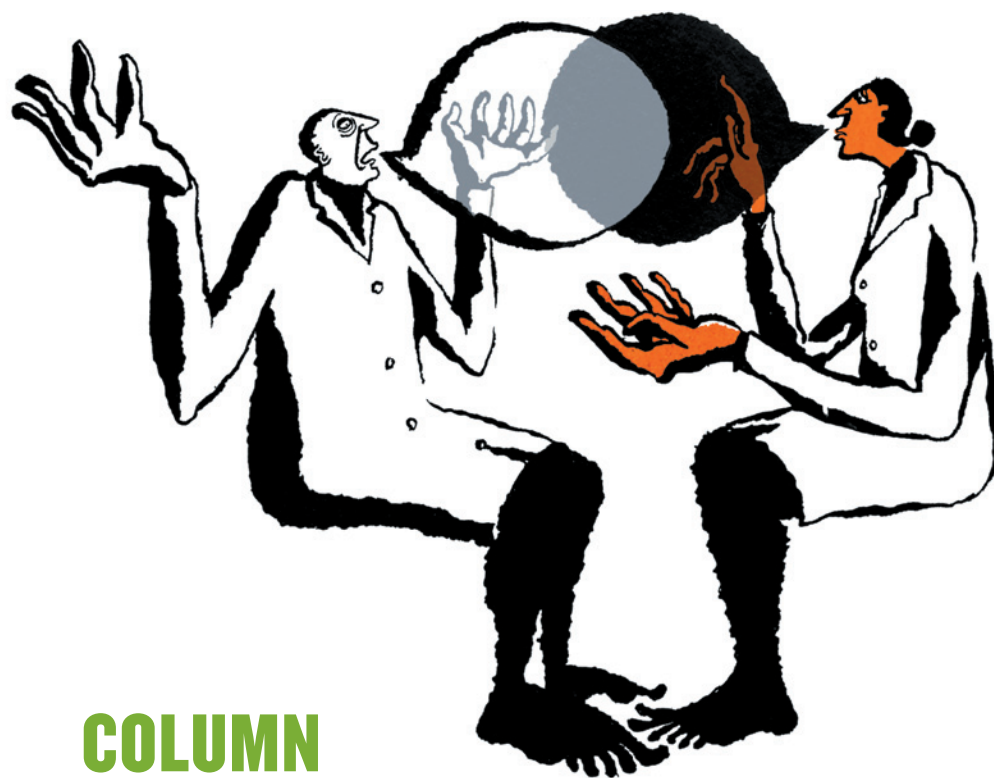
*The highest resolution shell is shown in parenthesis.

CAREERS

TURNING POINT Earth scientist unites disciplines to tackle social issues **p.429**

PUBLISHING Some retractions cause a decline in citations of previous work **p.429**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



COLUMN

The ethical grey zone

Confronting hypothetical dilemmas can ease workplace problems, argue **Caitlin Casey** and **Kartik Sheth**.

A colleague gets a nasty e-mail belittling her work. A student borrows data from a postdoc in his research group, not realizing that publishing it might constitute plagiarism. A researcher is being bullied, but his colleagues claim they are just kidding around and mean no harm. How should people witnessing such problems react?

Academia is rife with uncomfortable situations. To explore how researchers would respond to real-life murky dilemmas, we embarked on an in-person workshop and an online survey for astronomers. Participants ranked a range of scenarios on a continuum from desirable to unacceptable behaviour, without making stark judgements about right or wrong. The exercise made many participants uncomfortable, but it was eye-opening, raising awareness about issues such as bullying, harassment and unconscious biases that

currently plague our research community. Opening up a dialogue on these topics is the first step towards building a healthier research environment.

Scientists generally have much more training in analysing complex data sets than in how to handle potential ethical breaches or offensive comments in the workplace, whether inadvertent or intentional. We begin our research careers with the expectation that we and our colleagues will behave sensibly, appropriately and collaboratively. But in the competitive environment of the lab, the harsh reality of human nature sometimes surprises us.

The ethics and harassment training sessions that do exist prepare us for the most extreme inappropriate behaviours (outright threats, assault, weapons at work and quid pro quo harassment, in which, for example, a promotion is offered in exchange for sexual favours),

but they rarely address scenarios in the 'grey zone' — situations that might be unethical, undesirable or uncomfortable but are probably not severe enough to prompt legal action or reporting. How do we judge what is ethical and what is not? How should we react if we are uncomfortable with a colleague's behaviour?

CROWD-SOURCED ETHICS

At a workshop at the Aspen Center for Physics in Colorado in May, we were part of a group of astronomers who informally discussed how to build a positive, healthy work environment and make our community more inviting and inclusive of under-represented groups. We agreed that one major problem is lack of communication — from basic misunderstandings between colleagues all the way up to ignorance of academic work-environment protocols — and that one way to address these hurdles would be to get a large, diverse group of astronomers to discuss and rank some hypothetical scenarios.

In a subsequent session at the same meeting, we conducted a 'scenario-sorting' activity, in which astronomers were invited to discuss realistic situations involving the ethical ambiguities that our community faces every day: plagiarism, sexual harassment, hostile work environments, bullying, cultural clashes, unconscious biases and simple misunderstandings. Each scenario was printed on a slip of paper and handed to a participant, making sure that everyone had a different situation to contemplate.

We asked everyone to stand up and work together to organize their assigned scenarios, from the most desirable through acceptable, undesirable and unacceptable, to unethical. Once they had decided on the relative ranking, we discussed the scenarios as a group, exploring how participants with different backgrounds had made different judgements.

During group discussion, we often heard our colleagues exclaim in disbelief: "This cannot possibly be true!" The participants did not know that the 25 scenarios we had given them were not hypothetical — all came from firsthand experiences, whether our own or those of our colleagues, in the past 3–5 years. We had just changed names and revealing details to protect identities.

After we disclosed the truth, participants who had been sceptical about claims of harassment, hostility or plagiarism — including many senior male astronomers — admitted that the exercise was eye-opening and had ►

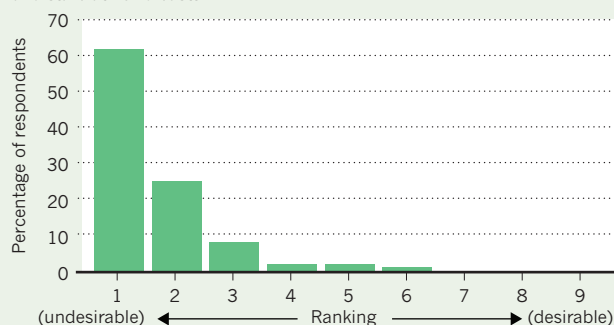
BELLE MELLOR

ETHICAL RANKINGS

In an online survey, astronomers reacted to various ethically uncomfortable scenarios. They converged on similar assessments of apparent sexual harassment (left) and plagiarism, but responses to scenarios involving more nuanced misunderstandings, stereotypes or unconscious biases were less clear-cut (right).

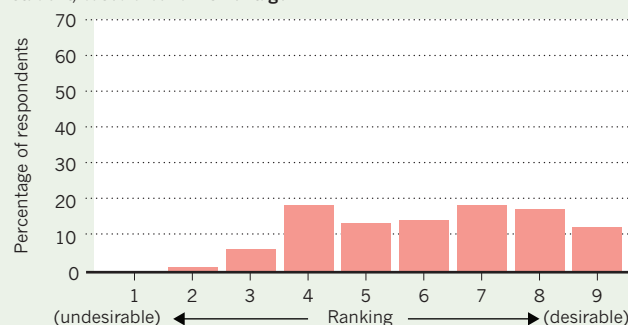
SCENARIO 1:

Older colleagues whistle at Janine and stare at her breasts.



SCENARIO 2:

A department throws a welcome party for Lucas, an international student, based around his heritage.



► forced them to think differently about their own interactions, especially with members of under-represented groups, including women and ethnic minorities.

After the success of the workshop, we took the exercise online through AstroBetter (www.astrobetter.com), a blog that aims to support the astronomy community. We used the same 25 scenarios and asked participants to rate each on a scale of 1–9, with 1 the worst and 9 the best. The response was overwhelming. Of the site's roughly 3,000 readers, 481 participated in the survey. More than 120 of those explained their thoughts on individual scenarios or on the survey overall.

GAUGING COMMUNITY STANDARDS

Some scenarios dealt with instances of academic ethical breaches such as apparent plagiarism, in which multiple characters and points of view made it difficult to determine the level of culpability. Others dealt with feelings or reactions rather than behaviours or actions. For example: "Brian was shortlisted for a faculty job, but the job went to a woman instead. Brian feels that it's unfair, because he thinks he would have got the job if he were a woman." We did not specify whose point of view the audience should analyse; the purpose was to trigger an emotional response for situations that some readers might not otherwise consider.

Scenarios generally did not telegraph a correct or 'appropriate' response. For example, one read: "Jane and John are new faculty members in a male-dominated department. Jane is told that she must serve on more faculty committees than John because they need a woman." Respondents might have considered, for example, how comfortable they were with Jane being instructed to take on more commitments because of her gender.

Participants agreed that sexual harassment is one of the more blatantly unethical practices in academia. However, we were particularly interested in the written feedback on a scenario in which a female astronomer is uncomfortable

wearing dresses to work because some senior professors whistle at her in the hallway and stare at her breasts. The collective judgement was that this scenario was one of the worst of the bunch (see 'Ethical rankings'). One online commenter said: "This makes me angry. And very sad."

However, another reader thought that there could be several levels of culpability. "I don't say that women should be blamed for whistling of men, but some clearly cross a line with too provocative outfits... whistling should be avoided but honest compliments on clothing style/appearance should be allowed." Another participant wrote that cultural context is important, noting that in conservative countries where women generally wear more clothing for religious or societal reasons, jeering at a woman wearing, for example, a skirt and high heels might be a socially acceptable response.

One reader declared adamantly that the researcher's actions or appearance are irrelevant and that "the professor's action is blatantly unethical no matter what the [researcher] is wearing". Indeed, in US universities, the scenario would be a textbook example of sexual harassment fostering a hostile work environment, and thus would be subject to legal action.

Some scenarios received a broad range of responses. In one, for example, a university department welcomes a new international student with a party celebrating his heritage. One participant said, "Making new people feel welcome is great, but singling out one's nationality while doing so seems a bit ham-handed." Another reader said it "depends entirely on how other incoming researchers/faculty/students are treated". Someone else commented: "This sounds nice at first, but it seems a little creepy or odd, if not presumptuous."

Two of the scenarios dealt with plagiarism. In one, a grant proposal for a large collaboration is plagiarized — the grant writer decides it is acceptable to copy her colleague's proposal because they are members of a single collaboration applying for funds. In the other, an idea is taken from a talk at a conference and published without appropriate credit. These scenarios received very strong feedback, all taking the same view. "No words. This is awful," said one respondent. Another said: "There's no way that the other guy can hijack someone's proposal; any co-investigator should stand up and protest."

REDEFINING RIGHT AND WRONG

Why is the community split on gender topics but not on plagiarism? Perhaps ambiguity in the descriptions left room for interpretation. Or perhaps there is a lack of community awareness about gender-based biases.

One online participant suggested a technique for checking for unconscious bias. "For all the gendered scenarios, I tried flipping the gender of each person in the situation and rereading it. This was an insightful exercise; several of my answers changed after the gender swap." The respondent acknowledges that unconscious gender biases influenced his or her answer. Realizing that unconscious biases might subtly steer our moral compasses is the first step towards abolishing them.

Scenarios that cannot be definitively classified as right or wrong can be intimidating, especially to those whose life's work is based on objective reason. But scientists in all fields can build a healthier work environment by considering their colleagues' disparate points of view — even if doing so means navigating ethical quandaries in decidedly grey areas. ■

Caitlin Casey is a McCue Postdoctoral Fellow of Cosmology at the University of California, Irvine, and **Kartik Sheth** is an astronomer at the US National Radio Astronomy Observatory in Charlottesville, Virginia.

TURNING POINT

Molly Brown

Molly Brown has spent a decade carving out a research field combining geography and economics. Using satellite images to trace how environmental factors affect food security, she helps government-aid agencies to pick priorities. On 29 October, Brown, an Earth scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, got an achievement award from the professional-development group Women in Aerospace.

How did your education shape your career?

I did a degree in biology at Tufts University in Medford, Massachusetts, and I did not like it very much. I hated memorizing names of things and found lab work isolating, but what bummed me out most was male teachers telling me about discoveries by great male scientists. With no mention of women, I almost couldn't see myself as a scientist. What got me excited was environmental science, an area in which many discoveries were yet to be made.

You also volunteered for the US Peace Corps. How did that set you on your career path?

I told the Peace Corps that they could send me anywhere they wanted. I ended up in Senegal, with few expectations. I come from a family of dairy farmers, but I knew nothing about subsistence farming. Over three years, I saw how the environment affects market dynamics. I watched grain rot in Senegal during a rainy, productive year because the region lacks a suitable transport system. A drought results in farmers losing their income source, and losing access to food as prices go up. When I got back to the United States, I decided to study geography because it combined my interests in people and the biophysical world.

What was your PhD research about?

While doing my PhD, I got a job at Goddard, helping to map a long-term remote-sensing record of when and where vegetation grows. The data were used by the Famine Early Warning Systems Network, a project of the US Agency for International Development (USAID) that informs distribution of monetary assistance. I decided to see if I could combine remote-sensing data with economic models to investigate how regional environmental conditions affect food-price dynamics, and to predict shocks to the markets.

How was that work received?

It made no impact whatsoever. I was creating a brand-new thing. Economists said I couldn't take this approach that looked beyond market



forces to satellite data. Geographers didn't understand it. It was very unpopular. A decade later I am finally getting traction.

How so?

In 2008 I published a book, *Famine Early Warning Systems and Remote Sensing Data* (Springer), so that I could place my innovation in the context of early-warning systems. My thesis had provided a method, but I needed to explain in detail how my approach could be used to predict crises. The book is much more widely cited than the academic publications from my thesis. I consider it a turning point. I have started working with high-calibre economists to move the idea forward, and I have another book due out in the spring.

How did the book help you to gain traction?

It let me explain who should care about my ideas and why it was essential to move them forward. You can do brilliant work, but if you are selling it to the wrong people, they won't buy it. I needed to target policy-makers, not just Earth scientists — I had to explain how the data could be used in decision-making. Now I am collaborating with USAID and the UN World Food Programme to propagate the idea of economists using spatial data.

What has been your best career strategy?

I approach academic pursuits as an entrepreneur. I am strategic about how to spend my time. I get the bulk of my funding from NASA, but I also develop research with collaborators to get extra grant funding. I have learned how to talk to people in a way that gets them to commit to projects. You have to be tough and claim ownership of your ideas to get people to fund them. ■

INTERVIEW BY VIRGINIA GEWIN

PUBLISHING

Retraction ripple effect

Journal-initiated retractions can reduce the number of citations of the author's earlier publications, a study finds (S. F. Lu *et al. Sci. Rep.* **3**, 3146; 2013). The authors analysed the effects of 667 retractions — mostly in the sciences and dating mainly from 2000 onwards — on citations of the author's earlier work. When a journal initiated the retraction, the number of annual citations of earlier papers fell by 6.9% on average. But author-initiated retractions had no such effect. The scientific community rewards honesty, says study co-author Ben Jones, an economist at Northwestern University in Evanston, Illinois. Self-reporting indicates that "you really care about getting it right", he says.

UNITED KINGDOM

Environment PhDs

The UK Natural Environment Research Council (NERC) is recruiting 1,200 PhD students in environmental sciences. Grants of about £82,000 (US\$130,000) will be paid over 3 or 4 years from 2014, as part of a £100-million government investment. Students will train with businesses, policy-makers or non-profit groups that might offer use of facilities or specialist training, or provide volunteer thesis examiners. One-third of recipients will conduct non-academic research, which could include working for an oil and gas company to find environmentally sustainable ways to extract resources, says Kirsty Grainger, NERC's head of skills and careers. "The student gets first-hand experience in the real world," she says.

UNITED STATES

Shutdown suffering

The 16-day US government shutdown that started on 1 October had serious effects on researchers, says a report by the US Office of Management and Budget (see go.nature.com/le3tcd). Some 98% of US National Science Foundation employees, two-thirds of employees of the US Centers for Disease Control and Prevention and three-quarters of US National Institutes of Health (NIH) employees were placed on mandatory leave, it says. Many NIH researchers could not enter their labs in Bethesda, Maryland. For early-career scientists, a few weeks' lost work "is a substantial percentage of their research experience", says Michael Gottesman, NIH deputy director for intramural research.

UNSOLVED LOGISTICAL PROBLEMS IN TIME TRAVEL: SPRING SEMESTER

Your time starts ... now.

BY MARISSA LINGEN

The student will offer a 2,000-word laboratory/field write-up solution to one of the following five problems. The student will deliver this solution to the instructor's account by midnight on 20 December, in order to give sufficient time for grading, instructor's subjective timeline. Any student demonstrably and successfully using a Marley device to obtain additional time for this final examination will be given 20 points extra credit on the course grade.

Students with alternative ideas about unsolved logistical problems in time travel should see the instructor for approval of these topics before (relative to their personal timeline) completing an alternate laboratory/field project. Points will be given for clarity, consistency and testability. Students who have not yet (relative to their personal timeline) completed TE1148: Mathematical Approaches to Time Travel Calibration, please note this in your file. All other students must demonstrate temporal consistency mathematically.

1. Cultural exchange with non-human predecessors/antecedents. Taking data on non-human reactions presents a series of challenges from observer effect to simple practicality. Write a plan for presenting a cultural artefact, e.g. Andrew Lloyd Webber's *Starlight Express*, to a non-human group, e.g. apatosaurs. Take into consideration factors such as the dietary needs of the predecessor/antecedent, such as apatosaurs' tendency to eat vegetative matter even when it forms a portion of the set.

2. Queuing theory for assassination tourism. If a dozen time travellers show up to assassinate Hitler in the chaos after the Beer Hall Putsch, who gets precedence? How do we fairly and practicably adjust the prioritization factors for era of origin, level of historical plausibility, use of worldline objects and other important factors? Show at least one example.

3. Asymptotic Grandfather Paradox refinements. Although it is well known that the laws of nature (cf. the Novikov self-consistency principle) intervene to keep scholars in the field of temporal dynamics from self-annihilation, exactly how close can one come to these issues? Novel ways of determining proximity are encouraged. Error bars will be

extremely important in the report of this field work. There is potential for publication as part of a larger project; please see instructor.

4. Forcing a branch point: parallel universes and the multiplicity of the absurd. Although travel between branch groups has hitherto been impossible (nor do we expect you to change that in a graduate-level course), what trivial changes are large enough to separate off branch groups? Special reference to the work of the Duchamp Theorists of Dada Time Travel is indicated, either to support or deprecate its claims. Students should perform very careful calculations and worldline object non-sentient dry runs to ensure ability to return to a timeline in which this course is held. Credit will be withheld in all timelines in which the course is not held or is taught by a different instructor. Credit will also be withheld from all solutions relying on a Gödel metric. Really, I shouldn't have to say this every term. You should have learned better in TE600.

5. Calculating destinations for small changes outside the main thrust of history. This project will require several examples for comparison and contrast, including early intervention in childhood, early intervention in career, and inspiration of late-period accomplished but obscure scientists, artists and other important figures. Extra credit for producing the kind of small change that ramifies in fields of the professor's particular interest, e.g. extrasolar planetary travel. The instructor does not mind suck-ups as long as they are competent suck-ups.

All travel must take place within the Visser–Roman Ring. Any student violating polygonal symmetry for the purposes of this course will be reported to the Dean of

➔ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

Students with serious repercussions including possible (likely!) expulsion. Symmetry is serious business.

Please note: Any student wishing to do an experiment falling under the sub-field 'Catalysing Effects of Lesser Dictatorships' will need to fill out *Form 753/J12: Experimentation On Human Populations* and demonstrate completion (relative to their personal timeline) of TE1120, Ethics Of Population Shock for Time Travellers.

Such experiments *must* have prior instructor approval and must be accompanied by an instructor or a teaching assistant at all times.

Students wishing to experiment on their own past selves must fill out *Form 753/J15: Waiver of Human Experimentation Protocol Forms*. Please bring with you at least three forms of proof that your past self is sufficiently congruent to your current self to qualify.

Students wishing to experiment on their own future selves should refer to the Student Counselling Services and obtain from them a *Form 753/J27: Waiver of Human Experimentation Protocol Forms — Special Case: Career Counselling*.

Office hours are Mondays 10–11 a.m., looped until all student concerns are addressed. If you notice more than four fellow students ahead of you, please bring the instructor a caffeinated beverage, because it's going to be a long day.

Spelling and grammar do matter. Inattention to detail can kill a time traveller. Check your own work. Then check your friends' work. Then check your enemies' work. Check your enemies' work again.

This is a laboratory/field course. All projects must have a travel component not only postulated but accomplished. Although modifications to spatial dampers and other essential equipment are welcome, they are outside the scope of this course, which is meant to be substantially a practicum. ■

Marissa Lingen has published more than 90 short stories in venues such as Analog, Jim Baen's Universe and Aeon Speculative Fiction.



JACEY

Three-dimensional imaging of dislocations

ARISING FROM C.-C. Chen *et al.* *Nature* **496**, 74–77 (2013)

At first sight, the achievement of determining atom positions in three dimensions appears spectacular¹. Chen and colleagues¹ apply a form of tomographic reconstruction to a tilt series of annular dark field (ADF) images of crystalline particles with defects, where the original data has a filter applied to reduce noise. However, the filtering imposes periodicities and significantly downgrades resolution, and the condition of signal linearity—a requirement for tomography—has not been met. We consider that their procedure gives an illusion of locating atom positions accurately. There is a Reply to this Brief Communication Arising by Miao, J. *et al.* *Nature* **503**, <http://dx.doi.org/10.1038/nature12661> (2013).

The experimental conditions (10.7 mrad beam convergence semi-angle at 200 keV) correspond to a resolution of about 0.23 nm, similar to that achieved in an earlier paper², where the overall shape of the particle was reconstructed. The most useful information from a lattice is obtained from projections down the various crystal zone axes, along which atoms line up into resolvable columns. As shown in earlier work^{3,4}, atomic resolution in ADF images is only possible when there is strong channelling down atomic columns. Simple geometry shows that a tilt of just 2° causes 7 nm columns, separated by 0.23 nm, to overlap each other in projection. Tilting from well-channelled conditions (zone axes) by only 2° generates significant intensity fluctuations, as shown in supplementary figure 1 of ref. 1, where bright and dark grains appear and disappear with tilt angle. This means that the key condition for tomography, that the recorded intensity be linearly related to projected potential, is violated.

When a channelled probe reaches a defect in a crystal, such as an edge dislocation, atomic columns become aligned with channels and vice versa. A well-channelled probe then encounters strong de-channelled conditions, resulting in loss of intensity. The effect is strongest in ADF images when the defect is close to the beam entrance surface^{5,6}.

Information on the position of atoms around a defect with long-range strain is in the diffuse scattering, not the lattice reflections^{7,8}. The procedure used in ref. 1 deliberately selects just the {200} and {111} Bragg peaks, and small regions around them (see figure 1b in ref. 1), suppressing essential diffuse contributions. In effect, this filter applies a point-spread function to each image, whose width is related to the inverse of the mask diameter around each reflection. This width, the resolution of the reconstruction, can be several times larger than the lattice spacing. It could be that essential diffuse scattering is lost in the noise, in which case this lowered resolution represents a fundamental limit on what can be achieved. Because the mask encompasses lattice reflections, the point-spread function is modulated by the lattice periodicity, giving the illusion of lattice resolution. Evidence for the downgraded resolution appears in supplementary figure 5c–e in ref. 1, which shows periodic ‘atomic columns’ outside the particle boundary. Although the procedure¹ might still locate dislocations and

grain boundaries, it does not necessarily put atoms in the right places because essential Fourier components are missing. This can be seen by comparing the atom positions in the rows immediately adjacent to the dislocation core in the model for the screw dislocation (supplementary figure 7b in ref. 1) and its reconstruction (supplementary figure 7d in ref. 1). Even when the noise threshold is set to 10%, there are still considerable displacements.

Some of the images presented¹ show Moiré fringes, indicating that the contrast is not just a simple linear projection of the atomic density of each layer. This is particularly evident in figure 3a in ref. 1. The layers are about two unit cells thick, and it is unlikely that they all contain an extended in-plane stacking fault. The depth resolution is apparently much larger (poorer) than the slice thickness, allowing significant mixing of information between slices.

We agree that the method presented by Chen *et al.*¹ identifies dislocations and defects in a crystal, but diffraction contrast alone can correctly identify the location and nature of dislocations⁶. The real challenge for tomography is to locate the 3D positions of all the atoms in an amorphous particle. We consider that the claims made¹ for the tomographic method on a crystalline particle are not appropriate; it is not a true 3D reconstruction giving precise atomic positions.

Peter Rez¹ & Michael M. J. Treacy¹

¹Department of Physics, Arizona State University, PO Box 871504, Tempe, Arizona 85287-1504, USA.
email: treacy@asu.edu

Received 14 May; accepted 11 September 2013.

1. Chen, C.-C. *et al.* Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution. *Nature* **496**, 74–77 (2013).
2. Scott, M. C. *et al.* Electron tomography at 2.4 Å resolution. *Nature* **483**, 444–447 (2012).
3. Jesson, D. E. & Pennycook, S. J. Incoherent imaging of crystals using thermally scattered electrons. *Proc. R. Soc. A* **449**, 273–293 (1995).
4. Hillyard, S., Loane, R. F. & Silcox, J. Annular dark-field imaging: resolution and thickness effects. *Ultramicroscopy* **49**, 14–25 (1993).
5. Treacy, M. M. J. & Gibson, J. M. in *Electron Microscopy and Analysis* 263–266 (Institute of Physics Conf. Ser. No. 61, 1981).
6. Hirsch, P. B., Howie, A., Nicholson, R. B., Pashley, D. W. & Whelan, M. J. *Electron Microscopy of Thin Crystals* 165–193; 222–275 (Butterworths, 1965).
7. Rez, P. *The Theory of Inelastic Scattering in the Electron Microscopy of Crystals*. D.Phil. thesis, Oxford Univ. (1976).
8. Anstis, G. R. & Cockayne, D. J. H. The calculation and interpretation of high-resolution electron microscope images of lattice defects. *Acta Crystallogr. A* **35**, 511–524 (1979).

Author Contributions The authors contributed equally.

Competing Financial Interests Declared none.

doi:10.1038/nature12660

Miao *et al.* reply

REPLYING TO P. Rez & M. M. J. Treacy *Nature* **503**, <http://dx.doi.org/10.1038/nature12660> (2013)

Although we welcome Rez and Treacy’s comment¹ on our paper², we find—on the basis of the considerations below—that these authors do not provide concrete scientific evidence to support their arguments,

and that their main statements are not consistent with our multi-slice simulations and experimental results using two independent filters.

First, it is well established in electron tomography that, if zone-axis orientations are avoided, images of a thin specimen obtained by high-angle annular dark-field scanning transmission electron microscopy (STEM) to a good approximation meet the projection requirement^{2–7}. We have confirmed this by performing extensive multislice STEM calculations combined with equally sloped tomography (EST) reconstructions^{2,3,8–10}. For those who are interested in verifying the results, both the multislice and EST software are available online (ref. 8 and <http://people.ccmr.cornell.edu/~kirkland>).

Second, in ref. 3, we not only reconstructed the three-dimensional (3D) surface morphology of a gold nanoparticle, but also revealed its internal lattice structure, identified several grains in three dimensions, and observed individual atoms in some regions of the nanoparticle.

Third, the discussion of Rez and Treacy about electron channelling and atomic resolution, “The most useful ... in projection.”¹ and “When a channelled ... entrance surface.”¹, is applicable to two-dimensional (2D) atomic-resolution imaging, but not to 3D atomic-resolution imaging with conventional electron tomography. In our tomography method, we achieved atomic-resolution 3D imaging of dislocations by avoiding electron channelling^{2,3}.

Fourth, besides multislice calculations, we have taken other measures to alleviate the nonlinear effects in the experimental data. For each tilt series, we project all projections onto the tilt axis to obtain a set of one-dimensional (1D) curves. If the projection requirement holds, all the 1D curves should be consistent. Furthermore, after obtaining an EST reconstruction, we calculate a set of projections from the 3D reconstruction at the same experimental tilt angles and compare them with the measured ones. Those inconsistent experimental projections are then removed^{2,3}. As for supplementary figure 1 in our paper², the bright/dark grains in several tilt angles are due to the existence of sub-grains, 3D surface morphology of the nanoparticle, and some diffraction-contrast in the images (supplementary video 1; ref. 2).

Fifth, if all we did was simple Fourier filtering with small apertures around the Bragg spots, then this would indeed lead to artefacts; we avoided this by verifying results against unbiased Wiener filters as well as by using relatively large apertures which were adjusted to minimize signal loss. The Wiener filter is well established for reducing the noise in a signal without any bias as to where the signal comes from^{11,12}. Supplementary figures 5 and 10 in ref. 2 show that the atomic positions obtained with 3D Fourier and Wiener filtering are consistent². The not-well-defined boundary in supplementary figure 5c–e (in ref. 2) is due to the convolution effect with Wiener and Fourier filtering. As for Supplementary Fig. 7d in ref. 2, although several atoms in the dislocation core are elongated (caused by noise, the missing wedge and a limited number of projections), the atomic positions agree with the model.

Sixth, the fuzziness in some parts of figure 3a (in ref. 2) is because the experimental Pt particle is a decahedral multiply-twinned nanoparticle, consisting of five main grains with different orientations².

When tilting the same 7.9-Å-thick slice to four different orientations, we see better lattice structure on the left, middle, right and bottom of the slice.

In conclusion, we have taken multiple measures to alleviate the nonlinear effects in our experimental data. We have also verified the 3D dislocation structures and 3D atomic positions in the reconstruction using three methods²: (1) multislice STEM calculations; (2) Wiener filtering; and (3) Fourier filtering. Carefully examining the atomic positions obtained by these independent methods suggests that the displacement due to 3D Fourier filtering is typically within one voxel along the *x*, *y* and *z* axes (voxel size = 0.53 Å). Finally, our recent numerical results indicate that our electron tomography method can be used to locate the positions of all the atoms in an amorphous particle¹³.

Jianwei Miao^{1,2}, Chien-Chun Chen^{1,2}, Chun Zhu^{1,2}, M. C. Scott^{1,2}, Edward R. White^{1,2}, Chin-Yi Chiu^{2,3}, B. C. Regan^{1,2}, Yu Huang^{2,3} & Laurence D. Marks⁴

¹Department of Physics and Astronomy, University of California, Los Angeles, California 90095, USA.
email: miao@physics.ucla.edu

²California NanoSystems Institute, University of California, Los Angeles, California 90095, USA.

³Department of Materials Science and Engineering, University of California, Los Angeles, California 90095, USA.

⁴Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60201, USA.

1. Rez, P. & Treacy, M. M. J. Three-dimensional imaging of dislocations. *Nature* **503**, <http://dx.doi.org/10.1038/nature12660> (2013).
2. Chen, C.-C. *et al.* Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution. *Nature* **496**, 74–77 (2013).
3. Scott, M. C. *et al.* Electron tomography at 2.4-ångström resolution. *Nature* **483**, 444–447 (2012).
4. Arslan, I., Yates, T. J. V., Browning, N. D. & Midgley, P. A. Embedded nanostructures revealed in three dimensions. *Science* **309**, 2195–2198 (2005).
5. Midgley, P. A. & Dunin-Borkowski, R. E. Electron tomography and holography in materials science. *Nature Mater.* **8**, 271–280 (2009).
6. Pennycook, S. J. & Nellist, P. D. *Scanning Transmission Electron Microscopy: Imaging and Analysis* 1st edn (Springer, 2011).
7. Van den Broek, W., Van Aert, S. & Van Dyck, D. A model based atomic resolution tomographic algorithm. *Ultramicroscopy* **109**, 1485–1490 (2009).
8. Equally Sloped Tomography (EST). <http://www.physics.ucla.edu/research/imaging/EST/> (2012).
9. Miao, J., Förster, F. & Levi, O. Equally sloped tomography with oversampling reconstruction. *Phys. Rev. B* **72**, 052103 (2005).
10. Lee, E. *et al.* Radiation dose reduction and image enhancement in biological imaging through equally sloped tomography. *J. Struct. Biol.* **164**, 221–227 (2008).
11. Brown, R. G. & Hwang, P. Y. C. *Introduction to Random Signals and Applied Kalman Filtering* 3rd edn (Wiley & Sons, 1996).
12. Marks, L. D. Wiener-filter enhancement of noisy HREM images. *Ultramicroscopy* **62**, 43–52 (1996).
13. Zhu, C. *et al.* Towards three-dimensional structural determination of amorphous materials at atomic resolution. *Phys. Rev. B* **88**, 100201 (2013).

doi:10.1038/nature12661